Applying Machine Learning Methods to Suggest Network Involvement and Functionality of Genes in Saccharomyces cerevisiae

By

Heather A. Amthauer

Submitted to the graduate degree program in Electrical Engineering and Computer Science and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

## Chairperson

Co-Chairperson

Committee members* $\qquad$ *
$\qquad$ *
$\qquad$ *
$\qquad$ *

Date defended: $\qquad$

The Dissertation Committee for Heather A. Amthauer certifies that this is the approved version of the following dissertation:

Applying Machine Learning Methods to Suggest Network Involvement and Functionality of Genes in Saccharomyces cerevisiae

Committee:

Chairperson*

Co-Chairperson

Date approved: $\qquad$

## Acknowledgements

I would like to thank Dr Costas Tsatsoulis for his extreme patience, guidance, and support he has given me through this process of me getting my Ph.D. He deserves more accolades than I can give. I would also like to thank the Madison \& Lila Self Graduate Fellowship for their support of my graduate work.

On a more personal note, I need to thank my husband, Matthew Smith, and my daughter, Ava Amthauer. They sacrificed a lot so I could complete this degree.


#### Abstract

Elucidating genetic networks provides the foundation for the development of new treatments or cures for diseased pathways, and determining novel gene functionality is critical for bringing a better understanding on how an organism functions as a whole. In this dissertation, I developed a methodology that correctly locates genes that may be involved in genetic networks with a given gene based on its location over $50 \%$ of the time or based on its description over $43 \%$ of the time. I also developed a methodology that makes it easier to predict how a gene product behaves in a cellular context by suggesting the correct Gene Ontology term over $80 \%$ of the time. The designed software provides researchers with a way to focus their search for coregulated genes which will lead to better microarray chip design and limits the list of possible functions of a gene product. This ultimately saves the researcher time and money.


## Table of Contents

1. Introduction .....  1
1.1 Significance of Problem ..... 2
1.2 Other Approaches ..... 4
1.3 My Approach ..... 6
1.4 Contributions ..... 8
2. Background ..... 10
2.1 Knowledge Sources ..... 10
2.1.1 Saccharomyces Genome Database ..... 10
2.1.2 Gene Ontology ..... 10
2.1.3 KEGG ..... 11
2.2 Learning ..... 11
2.2.1 Classification Learning ..... 12
2.2.1.1 Bayesian Classifiers ..... 13
2.2.1.2 Tree Classifiers ..... 13
2.2.1.3 Functions ..... 14
2.2.1.4 Lazy Classifiers ..... 14
2.2.2 Classification Algorithms ..... 15
2.2.2.1 Naïve Bayes ..... 15
2.2.2.2 MultiBoostAB using the J48 classifier ..... 17
2.2.2.3 KStar ..... 22
2.3 Information Retrieval ..... 23
2.4 Patterns in Genomes ..... 24
2.5 Possible explanation of clusters/groups of genes in genomes ..... 30
2.6 Current Approaches ..... 33
2.6.1 Determining Networks ..... 33
2.6.2 Determining Functionality ..... 38
2.6.3 Establishing Similarity Using Text ..... 39
3. Methodology ..... 42
3.1 Approach. ..... 42
3.2 Systems ..... 45
4. Experiments ..... 50
4.1 Descriptions of Experiments ..... 50
4.1.1. Classifying Features to Networks ..... 50
4.1.2. Classifying Features to Networks ..... 51
4.1.3 Suggesting Locations of Interest ..... 52
4.2 Datasets ..... 53
4.3 Algorithms ..... 59
4.4 Metrics ..... 60
5. Results ..... 65
5.1 Classifying Genomic Features to Networks ..... 65
5.1.1 Using Neighbor's SGDID. .....  .65
5.1.1.1 Results for the Entire Genome ..... 65
5.1.1.2 Results for Individual Chromosomes ..... 74
5.1.1.2.1 Percent Correctly Classified. ..... 74
5.1.1.2.2 F-measures ..... 88
5.1.2 Using Neighbor's Gene Ontology Term ..... 100
5.1.2.1 Results for the Entire Genome ..... 100
5.1.2.2 Result for Individual Chromosomes ..... 110
5.1.2.2.1 Percent Correctly Classified. ..... 110
5.1.2.2.2 F-measures ..... 124
5.1.2.3 Removing Location Information ..... 135
5.1.3 Using Neighbor's Network Involvement. ..... 137
5.1.3.1 Results for the Entire Genome ..... 138
5.1.3.2 Result for Individual Chromosomes ..... 148
5.1.3.2.1 Percent Correctly Classified. ..... 148
5.1.3.2.2 F-measures. ..... 162
5.2 Classifying Genomic Features to Gene Ontology Terms ..... 173
5.2.1 Using Neighbor's SGDID. ..... 173
5.2.1.1 Results for the Entire Genome ..... 173
5.2.1.2 Result for Individual Chromosomes ..... 181
5.2.1.2.1 Percent Correctly Classified. ..... 181
5.2.1.2.2 F-measures ..... 198
5.2.2 Using Neighbor's Gene Ontology Information. ..... 211
5.2.2.1 Results for the Entire Genome. ..... 212
5.2.2.2 Result for Individual Chromosomes ..... 222
5.2.2.2.1 Percent Correctly Classified. ..... 222
5.2.2.2.2 F-measures ..... 236
5.2.2.3 Removing Location Information. ..... 252
5.3 Suggesting Locations...... 254
5.3.1 Classifying Genomic Features to Chromosome ..... 254
5.3.1.1 Percent of Correctly Classified. ..... 255
5.3.1.2 Effect on Individual Chromosomes... ..... 255
5.3.2 Classifying Genomic Features to Start Position Using Network Involvement. ..... 256
5.3.2.1 Generating Rules to Classify Start Position Using Network Involvement Information ..... 259
5.3.3 Classifying Genomic Features to Start Position Using Gene Ontology Term ..... 259
5.3.4 Using Genomic Feature Descriptions and Information Retrieval. ..... 261
5.4 Summary ..... 262
6. Discussion. ..... 265
6.1 Discussion of Results. ..... 265
6.2 Future Studies. ..... 267
7. References ..... 270
Appendix A: Stop List Used in Information Retrieval ..... 276
Appendix B: F-measures for the networks using KStar classifier. ..... 277
Appendix C: F-measures for the networks using the MultiBoostAB algorithm using the J48 classifier ..... 281
Appendix D: F-measures for the networks using KStar classifier ..... 285
Appendix E: F-measures for the GO term using the MultiBoostAB algorithm using the J48 classifier ..... 289
Appendix F: F-measures for the GO term using the MultiBoostAB algorithm using the J48 classifier ..... 291
Appendix G: Nine Rule Adaption. ..... 293
Appendix H: Decision Trees Generated for Classifying Features to Networks ..... 295
Appendix I: Decision Trees Generated for Classifying Features to Gene Ontology Terms ..... 449

## List of Tables

4.1 File Format for Network Involvement Classification with No Neighbors ..... 54
4.2 File Format for Network Involvement Classification Including Neighbors' SGDID's ..... 55
4.3 File Format for Network Involvement Classification Including Neighbors' Gene Ontology Terms ..... 55
4.4 File Format for Network Involvement Classification Including Neighbors' Network Involvement. ..... 56
4.5 File Format for Gene Ontology Term Classification with No Neighbors ..... 56
4.6 File Format for Gene Ontology Term Classification Including Neighbors' SGDID's ..... 57
4.7 File Format for Gene Ontology Term Classification Including Neighbors' Gene Ontology Information. ..... 58
4.8 Interpreting Kappa ..... 61
4.9 Interpreting Area Under ROC Curve. ..... 64
5.1 Percentage of Correctly Classified Features to Start Position Given Chromosome Number, Network Involvement and Start Position to Train on for the Entire Genome. ..... 256
5.2 A Representation of the Rules Generated by the MultiBoostAB Algorithm Using the J48 Classifier ..... 257
5.3 Percentage of Correctly Classified Features to Start Position Given Chromosome Number, Network Involvement and Start Position to Train on for the Individual Chromosomes ..... 258
5.4 Percentage of Correctly Classified Features to Start Position Given Network Involvement and Start Position to Train on for the Entire Genome ..... 258
5.5 Percentage of Correctly Classified Features to Start Position Given Chromosome Number, Network Involvement and Start Position to Train on for the Entire Genome.......... 259
5.6 Percentage of Correctly Classified Features to Start Position Given Gene Ontology Termand Start Position to Train on for the Individual Chromosomes.260
5.7 Percentage of Correctly Classified Features to Start Position Given Gene Ontology Term and Start Information to Train on for Entire Genome. ..... 261
5.8 Summary of the Retrieval Results and Network Suggestions Genomic Features. ..... 262
B1 F-measures for the Networks Using KStar Classifier ..... 277
C1 F-measures for the Networks Using the MultiBoostAB Algorithm Using the J48 Classifier ..... 281
D1 F-measures for the Networks Using KStar Classifier. ..... 285
E1 F-measures for the Gene Ontology Term using the MultiBoostAB Algorithm Using the J48 Classifier. ..... 289
F1 F-measures for the Gene Ontology Term Using the MultiBoostAB Algorithm Using the J48 Classifier ..... 291

## List of Figures

2.1 From Caron et al. Regional Expression Profiles for 23 Human Chromosomes ..... 27
2.2 Incidental Regulation. ..... 31
2.3 A Structural Domain Model ..... 32
3.1. Diagram of Software Designed to Suggest Areas of Interest for Networks...... ..... 46
3.2. Diagram of Software Designed to Suggest Gene Ontology Terms ..... 47
3.3. Diagram of How Location Suggestion is Made ..... 48
4.1. Illustrates the Interpretation of the Area Under the ROC Curve. ..... 65
5.1 Percent of Genomic Features Classified for the Entire Genome Using Neighbor SGDID to Classify Features to Networks ..... 66
5.2 Kappa Statistic for the Entire Genome Using Neighbor SGDID to Classify Features to Networks ..... 67
5.3 Precision for the Entire Genome Using Neighbor SGDID to Classify Features to Networks ..... 67
5.4 Recall for the Entire Genome Using Neighbor SGDID to Classify Features to Networks ..... 68
5.5 F-measure for the Entire Genome Using Neighbor SGDID to Classify Features to Networks ..... 69
5.6 Area Under ROC curve for the Entire Genome Using Neighbor SGDID to Classify
Features to Networks .....  70
5.7 F-measures for Networks Using Neighbor SGDID to Classify Features to Networks. ..... 71
5.8 F-measures for Networks Using Neighbor SGDID to Classify Features to Networks... ..... 72
5.9 F-measures for Networks Using Neighbor SGDID to Classify Features to Networks. ..... 73
5.10 Percent of Genomic Features Correctly Classified for Chromosome One Using Neighbor SGDID to Classify Features to Networks ..... 75
5.11 Percent of Genomic Features Correctly Classified for Chromosome Two Using Neighbor SGDID to Classify Features to Networks ..... 76
5.12 Percent of Genomic Features Correctly Classified for Chromosome Three Using Neighbor SGDID to Classify Features to Networks ..... 76
5.13 Percent of Genomic Features Correctly Classified for Chromosome Four Using Neighbor SGDID to Classify Features to Networks ..... 77
5.14 Percent of Genomic Features Correctly Classified for Chromosome Five Using Neighbor SGDID to Classify Features to Networks ..... 78
5.15 Percent of Genomic Features Correctly Classified for Chromosome Six Using Neighbor SGDID to Classify Features to Networks ..... 78
5.16 Percent of Genomic Features Correctly Classified for Chromosome Seven Using Neighbor SGDID to Classify Features to Networks ..... 79
5.17 Percent of Genomic Features Correctly Classified for Chromosome Eight Using Neighbor SGDID to Classify Features to Networks ..... 80
5.18 Percent of Genomic Features Correctly Classified for Chromosome Nine Using Neighbor SGDID to Classify Features to Networks ..... 81
5.19 Percent of Genomic Features Correctly Classified for Chromosome Ten Using Neighbor SGDID to Classify Features to Networks ..... 82
5.20 Percent of Genomic Features Correctly Classified for Chromosome Eleven Using Neighbor SGDID to Classify Features to Networks ..... 83
5.21 Percent of Genomic Features Correctly Classified for Chromosome Twelve Using Neighbor SGDID to Classify Features to Networks ..... 84
5.22 Percent of Genomic Features Correctly Classified for Chromosome Thirteen Using
Neighbor SGDID to Classify Features to Networks ..... 85
5.23 Percent of Genomic Features Correctly Classified for Chromosome Fourteen Using Neighbor SGDID to Classify Features to Networks ..... 86
5.24 Percent of Genomic Features Correctly Classified for Chromosome Fifteen Using Neighbor SGDID to Classify Features to Networks ..... 86
5.25 Percent of Genomic Features Correctly Classified for Chromosome Sixteen Using Neighbor SGDID to Classify Features to Networks ..... 87
5.26 F-measures for Chromosome One Using Neighbor SGDID to Classify Features to Networks ..... 88
5.27 F-measures for Chromosome Two Using Neighbor SGDID to Classify Features to Networks ..... 89
5.28 F-measures for Chromosome Three Using Neighbor SGDID to Classify Features to Networks ..... 90
5.29 F-measures for Chromosome Four Using Neighbor SGDID to Classify Features to Networks ..... 91
5.30 F-measures for Chromosome Five Using Neighbor SGDID to Classify Features to Networks ..... 91
5.31 F-measures for Chromosome Six Using Neighbor SGDID to Classify Features to Networks .....  92
5.32 F-measures for Chromosome Seven Using Neighbor SGDID to Classify Features to Networks ..... 93
5.33 F-measures for Chromosome Eight Using Neighbor SGDID to Classify Features to Networks ..... 93
5.34 F-measures for Chromosome Nine Using Neighbor SGDID to Classify Features to Networks ..... 94
5.35 F-measures for Chromosome Ten Using Neighbor SGDID to Classify Features to Networks ..... 95
5.36 F-measures for Chromosome Eleven Using Neighbor SGDID to Classify Features to Networks ..... 96
5.37 F-measures for Chromosome Twelve Using Neighbor SGDID to Classify Features to Networks ..... 97
5.38 F-measures for Chromosome Thirteen Using Neighbor SGDID to Classify Features to Networks ..... 97
5.39 F-measures for Chromosome Fourteen Using Neighbor SGDID to Classify Features to Networks ..... 98
5.40 F-measures for Chromosome Fifteen Using Neighbor SGDID to Classify Features to Networks ..... 99
5.41 F-measures for Chromosome Sixteen Using Neighbor SGDID to Classify Features to Networks ..... 99
5.42 Percent of Genomic Features Classified for the Entire Genome Using Neighbor Gene Ontology Terms to Classify Features to Networks. ..... 101
5.43 Kappa Statistic for the Entire Genome Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 102
5.44 Precision for the Entire Genome Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 103
5.45 Recall for the Entire Genome Using Neighbor Gene Ontology Terms to Classify Features to Networks. ..... 104
5.46 F-measure for the Entire Genome Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 105
5.47 Area Under ROC curve for the Entire Genome Using Neighbor Gene Ontology Terms To Classify Features to Networks ..... 106
5.48 F-measures for Networks Using Neighbor Gene Ontology Terms to Classify Features to Networks. ..... 107
5.49 F-measures for Networks Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 108
5.50 F-measures for Networks Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 109
5.51 Percent of Genomic Features Correctly Classified for Chromosome One Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 111
5.52 Percent of Genomic Features Correctly Classified for Chromosome Two Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 112
5.53 Percent of Genomic Features Correctly Classified for Chromosome Three Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 112
5.54 Percent of Genomic Features Correctly Classified for Chromosome Four Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 113
5.55Percent of Genomic Features Correctly Classified for Chromosome Five Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 114
5.56 Percent of Genomic Features Correctly Classified for Chromosome Six Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 115
5.57 Percent of Genomic Features Correctly Classified for Chromosome Seven Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 116
5.58 Percent of Genomic Features Correctly Classified for Chromosome Eight Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 117
5.59 Percent of Genomic Features Correctly Classified for Chromosome Nine Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 118
5.60 Percent of Genomic Features Correctly Classified for Chromosome Ten Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 119
5.61 Percent of Genomic Features Correctly Classified for Chromosome Eleven Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 119
5.62 Percent of Genomic Features Correctly Classified for Chromosome Twelve Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 120
5.63 Percent of Genomic Features Correctly Classified for Chromosome Thirteen Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 121
5.64 Percent of Genomic Features Correctly Classified for Chromosome Fourteen Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 122
5.65 Percent of Genomic Features Correctly Classified for Chromosome Fifteen Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 122
5.66 Percent of Genomic Features Correctly Classified for Chromosome Sixteen Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 123
5.67 F-measures for Chromosome One Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 124
5.68 F-measures for Chromosome Two Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 125
5.69 F-measures for Chromosome Three Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 126
5.70 F-measures for Chromosome Four Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 127
5.71 F-measures for Chromosome Five Using Neighbor Gene Ontology Terms to Classify
Features to Networks ..... 127
5.72 F-measures for Chromosome Six Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 128
5.73 F-measures for Chromosome Seven Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 128
5.74 F-measures for Chromosome Eight Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 129
5.75 F-measures for Chromosome Nine Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 130
5.76 F-measures for Chromosome Ten Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 131
5.77 F-measures for Chromosome Eleven Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 131
5.78 F-measures for Chromosome Twelve Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 132
5.79 F-measures for Chromosome Thirteen Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 133
5.80 F-measures for Chromosome Fourteen Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 133
5.81 F-measures for Chromosome Fifteen Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 134
5.82 F-measures for Chromosome Sixteen Using Neighbor Gene Ontology Terms to Classify Features to Networks ..... 135
5.83 Percent of Correctly Classified Features for the Entire Genome Using Neighbor Gene Ontology Terms and No Location to Classify Features to Networks. ..... 136
5.84 F-measure for the Entire Genome Using Neighbor Gene Ontology Terms and No Location to Classify Features to Networks. ..... 137
5.85 Percent of Genomic Features Classified for the Entire Genome Using Neighbor Network Involvement to Classify Features to Networks ..... 138
5.86 Kappa Statistic for the Entire Genome Using Neighbor Network Involvement to Classify Features to Networks ..... 139
5.87 Precision for the Entire Genome Using Neighbor Network Involvement to Classify Features to Networks ..... 140
5.88 Recall for the Entire Genome Using Neighbor Network Involvement to Classify Features to Networks. ..... 141
5.89 F-measure for the Entire Genome Using Neighbor Network Involvement to Classify Features to Networks. ..... 142
5.90 Area Under ROC curve for the Entire Genome Using Neighbor Network Involvement to Classify Features to Networks ..... 143
5.91 F-measures for Networks Using Neighbor Network Involvement to Classify Features to Networks ..... 145
5.92 F-measures for Networks Using Neighbor Network Involvement to Classify Features to Networks ..... 146
5.93 F-measures for Networks Using Neighbor Network Involvement to Classify Features to Networks ..... 147
5.94 Percent of Genomic Features Correctly Classified for Chromosome One Using Neighbor Network Involvement to Classify Features to Networks ..... 149
5.95 Percent of Genomic Features Correctly Classified for Chromosome Two Using Neighbor Network Involvement to Classify Features to Networks ..... 150
5.96 Percent of Genomic Features Correctly Classified for Chromosome Three Using Neighbor Network Involvement to Classify Features to Networks ..... 150
5.97 Percent of Genomic Features Correctly Classified for Chromosome Four Using Neighbor Network Involvement to Classify Features to Networks ..... 151
5.98 Percent of Genomic Features Correctly Classified for Chromosome Five Using Neighbor Network Involvement to Classify Features to Networks ..... 152
5.99 Percent of Genomic Features Correctly Classified for Chromosome Six Using Neighbor Network Involvement to Classify Features to Networks ..... 153
5.100 Percent of Genomic Features Correctly Classified for Chromosome Seven Using Neighbor Network Involvement to Classify Features to Networks ..... 154
5.101 Percent of Genomic Features Correctly Classified for Chromosome Eight Using Neighbor Network Involvement to Classify Features to Networks ..... 155
5.102 Percent of Genomic Features Correctly Classified for Chromosome Nine Using Neighbor Network Involvement to Classify Features to Networks ..... 156
5.103 Percent of Genomic Features Correctly Classified for Chromosome Ten Using Neighbor Network Involvement to Classify Features to Networks ..... 156
5.104 Percent of Genomic Features Correctly Classified for Chromosome Eleven Using Neighbor Network Involvement to Classify Features to Networks ..... 157
5.105 Percent of Genomic Features Correctly Classified for Chromosome Twelve Using Neighbor Network Involvement to Classify Features to Networks ..... 158
5.106 Percent of Genomic Features Correctly Classified for Chromosome Thirteen Using Neighbor Network Involvement to Classify Features to Networks ..... 159
5.107 Percent of Genomic Features Correctly Classified for Chromosome Fourteen Using Neighbor Network Involvement to Classify Features to Networks. ..... 160
5.108 Percent of Genomic Features Correctly Classified for Chromosome Fifteen Using Neighbor Network Involvement to Classify Features to Networks ..... 160
5.109 Percent of Genomic Features Correctly Classified for Chromosome Sixteen Using Neighbor Network Involvement to Classify Features to Networks ..... 161
5.110 F-measures for Chromosome One Using Neighbor Network Involvement to Classify Features to Networks ..... 162
5.111F-measures for Chromosome Using Neighbor Network Involvement to Classify Features to Networks ..... 163
5.112 F-measures for Chromosome Three Using Neighbor Network Involvement to Classify Features to Networks. ..... 164
5.113 F-measures for Chromosome Four Using Neighbor Network Involvement to Classify Features to Networks ..... 164
5.114 F-measures for Chromosome Five Using Neighbor Network Involvement to Classify Features to Networks ..... 165
5.115 F-measures for Chromosome Six Using Neighbor Network Involvement to Classify Features to Networks ..... 165
5.116 F-measures for Chromosome Seven Using Neighbor Network Involvement to Classify Features to Networks ..... 166
5.117 F-measures for Chromosome Eight Using Neighbor Network Involvement to Classify Features to Networks ..... 167
5.118 F-measures for Chromosome Nine Using Neighbor Network Involvement to Classify Features to Networks ..... 167
5.119 F-measures for Chromosome Ten Using Neighbor Network Involvement to Classify Features to Networks ..... 168
5.120 F-measures for Chromosome Eleven Using Neighbor Network Involvement to
Classify Features to Networks ..... 169
5.121 F-measures for Chromosome Twelve Using Neighbor Network Involvement to Classify Features to Networks ..... 170
5.122 F-measures for Chromosome Thirteen Using Neighbor Network Involvement to Classify Features to Networks ..... 170
5.123 F-measures for Chromosome Fourteen Using Neighbor Network Involvement to Classify Features to Networks ..... 171
5.124 F-measures for Chromosome Fifteen Using Neighbor Network Involvement to Classify Features to Networks ..... 172
5.125 F-measures for Chromosome Sixteen Using Neighbor Network Involvement to Classify Features to Networks ..... 172
5.126 Percent of Genomic Features Classified for the Entire Genome Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 174
5.127 Kappa Statistic for the Entire Genome Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 175
5.128 Precision for the Entire Genome Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 175
5.129 Recall for the Entire Genome Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 176
5.130 F-measure for the Entire Genome Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 177
5.131 Area Under ROC curve for the Entire Genome Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 178
5.132 F-measures for Gene Ontology Terms Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 179
5.133 F-measures for Gene Ontology Terms Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 180
5.134 Percent of Genomic Features Correctly Classified for Chromosome One Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 182
5.135 Percent of Genomic Features Correctly Classified for Chromosome Two Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 183
5.136 Percent of Genomic Features Correctly Classified for Chromosome Three Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 184
5.137 Percent of Genomic Features Correctly Classified for Chromosome Four Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 185
5.138 Percent of Genomic Features Correctly Classified for Chromosome Five Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 186
5.139 Percent of Genomic Features Correctly Classified for Chromosome Six Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 187
5.140 Percent of Genomic Features Correctly Classified for Chromosome Seven Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 188
5.141 Percent of Genomic Features Correctly Classified for Chromosome Eight Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 189
5.142 Percent of Genomic Features Correctly Classified for Chromosome Nine Using Neighbor SGDID to Classify Features to Gene Ontology Term. ..... 190
5.143 Percent of Genomic Features Correctly Classified for Chromosome Ten Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 191
5.144 Percent of Genomic Features Correctly Classified for Chromosome Eleven Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 192
5.145 Percent of Genomic Features Correctly Classified for Chromosome Twelve Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 193
5.146 Percent of Genomic Features Correctly Classified for Chromosome Thirteen Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 194
5.147 Percent of Genomic Features Correctly Classified for Chromosome Fourteen Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 195
5.148 Percent of Genomic Features Correctly Classified for Chromosome Fifteen Using Neighbor SGDID to Classify Features to Gene Ontology Term. ..... 196
5.149 Percent of Genomic Features Correctly Classified for Chromosome Sixteen Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 197
5.150 Percent of Genomic Features Correctly Classified for Mitochondrial Chromosome Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 197
5.151 F-measures for Chromosome One Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 199
5.152 F-measures for Chromosome Two Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 200
5.153 F-measures for Chromosome Three Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 201
5.154 F-measures for Chromosome Four Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 201
5.155 F-measures for Chromosome Five Using Neighbor SGDID to Classify Features to Gene Ontology Term .....  202
5.156 F-measures for Chromosome Six Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 203
5.157 F-measures for Chromosome Seven Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 203
5.158 F-measures for Chromosome Eight Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 204
5.159 F-measures for Chromosome Nine Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 205
5.160 F-measures for Chromosome Ten Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 206
5.161 F-measures for Chromosome Eleven Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 206
5.162 F-measures for Chromosome Twelve Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 207
5.163 F-measures for Chromosome Thirteen Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 208
5.164 F-measures for Chromosome Fourteen Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 209
5.165 F-measures for Chromosome Fifteen Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 209
5.166 F-measures for Chromosome Sixteen Using Neighbor SGDID to Classify Features to Gene Ontology Term. ..... 210
5.167 F-measures for Mitochondrial Chromosome Using Neighbor SGDID to Classify Features to Gene Ontology Term ..... 211
5.168 Percent of Genomic Features Classified for the Entire Genome Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 213
5.169 Kappa Statistic for the Entire Genome Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 214
5.170 Precision for the Entire Genome Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 215
5.171 Recall for the Entire Genome Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 216
5.172 F-measure for the Entire Genome Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 217
5.173 Area Under ROC curve for the Entire Genome Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 218
5.174 F-measures for Gene Ontology Terms Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 220
5.175 F-measures for Gene Ontology Terms Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 221
5.176 Percent of Genomic Features Correctly Classified for Chromosome One Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 222
5.177 Percent of Genomic Features Correctly Classified for Chromosome Two Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 223
5.178 Percent of Genomic Features Correctly Classified for Chromosome Three Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term. ..... 224
5.179 Percent of Genomic Features Correctly Classified for Chromosome Four Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 225
5.180 Percent of Genomic Features Correctly Classified for Chromosome Five Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term. ..... 226
5.181 Percent of Genomic Features Correctly Classified for Chromosome Six Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 226
5.182 Percent of Genomic Features Correctly Classified for Chromosome Seven Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 227
5.183 Percent of Genomic Features Correctly Classified for Chromosome Eight Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 228
5.184 Percent of Genomic Features Correctly Classified for Chromosome Nine Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 228
5.185 Percent of Genomic Features Correctly Classified for Chromosome Ten Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 229
5.186 Percent of Genomic Features Correctly Classified for Chromosome Eleven Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 230
5.187 Percent of Genomic Features Correctly Classified for Chromosome Twelve Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 231
5.188 Percent of Genomic Features Correctly Classified for Chromosome Thirteen Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 232
5.189 Percent of Genomic Features Correctly Classified for Chromosome Fourteen Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 233
5.190 Percent of Genomic Features Correctly Classified for Chromosome Fifteen Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 234
5.191 Percent of Genomic Features Correctly Classified for Chromosome Sixteen Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 235
5.192 Percent of Genomic Features Correctly Classified for Mitochondrial Chromosome Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 236
5.193 F-measures for Chromosome One Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 237
5.194 F-measures for Chromosome Two Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 238
5.195 F-measures for Chromosome Three Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 238
5.196 F-measures for Chromosome Four Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 239
5.197 F-measures for Chromosome Five Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 240
5.198 F-measures for Chromosome Six Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 241
5.199 F-measures for Chromosome Seven Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 242
5.200 F-measures for Chromosome Eight Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 243
5.201 F-measures for Chromosome Nine Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 244
5.202 F-measures for Chromosome Ten Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 245
5.203 F-measures for Chromosome Eleven Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 246
5.204 F-measures for Chromosome Twelve Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 247
5.205 F-measures for Chromosome Thirteen Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 248
5.206 F-measures for Chromosome Fourteen Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 249
5.207 F-measures for Chromosome Fifteen Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 250
5.208 F-measures for Chromosome Sixteen Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 251
5.209 F-measures for Mitochondrial Chromosome Using Neighbor Gene Ontology Information to Classify Features to Gene Ontology Term ..... 252
5.210 Percent of Correctly Classified Features for the Entire Genome Using Neighbor Gene Ontology Information and No Location to Classify Features to Gene Ontology Term ..... 253
5.211 F-measure for the Entire Genome Using Neighbor Gene Ontology Information and No Location to Classify Features to Gene Ontology Term ..... 254
5.212 Percent of Correctly Classified Features for the Entire Genome to Classify Features to Chromosome ..... 255
5.213 Percent of Correctly Classified Features for the Chromosomes to Classify Features to Chromosome. ..... 256

## 1 Introduction

Several studies have challenged the notion that gene location in higher eukaryotic organisms is random. These studies [Blumenthal, 1998; Boutanaev et al., 2002; Caron et al., 2001; Cohen et al., 2000; Lee \& Sonnhammer, 2003; Lercher et al., 2002; Li et al., 2005; Mayor et al., 2004; Roy et al. 2002; Spellman \& Rubin, 2002] suggest that there may be patterns in gene location. After reading these studies, it became obvious that the insight that data mining has to offer was not being taken advantage of. Little work is being dedicated to determining if there are true patterns within a genome, and if we can learn from these patterns.

As we all know, in most real estate markets, the location of a house and information about the surrounding houses greatly influences its value. If a genome is analogous to a collection of neighborhoods and genomic features are analogous to houses in those neighborhoods, then this type of phenomenon has also been found in the yeast genome in my research. Knowing the location of a genomic feature and information about that feature's nearest neighbors gives us insight into what genetic network the feature belongs to and how its gene product(s) behave in a cellular context (Gene Ontology term).

In this dissertation, I developed a methodology to help locate genes that may be involved in genetic networks with a given gene based on its location or its description. This is an important concept because by making it easier to locate other genes that interact with a given gene, we are able to learn more about that given gene by seeing how it influences other genes. This is very critical in elucidating pathways because it may allow for pathway initiators (which are ideal targets for drug therapy) to be located faster. Being able to elucidate genetic networks provides the
foundation on which new treatments or cures for diseased pathways to be developed. We can design drugs to target specific genes in the flawed networks.

I also developed a methodology to makes it easier to predict how a gene product behaves in a cellular context. Determining novel gene functionality is critical in bringing a better understanding on how an organism functions as a whole. Lowering the number of potential functionalities for a given gene product can make the process of identifying the true functionality faster.

My experimental results showed that my methodology is able to correctly locate areas of interest for network involvement over $50 \%$ of the time and over $80 \%$ of the time suggest the correct Gene Ontology term(s) for a given genomic feature. The major contribution of my work is in the fields of biology and bioinformatics, because it allows scientists to use what they already know to learn something new about a given genomic feature. By analyzing data from a different perspective (i.e. focusing on the location and the contribution of the nearest neighbors of each genomic feature), I was able to generate rules that can be used to focus searches for genetic networks or determine gene product functionality.

### 1.1 Significance of Problem

Discovering genetic networks is important. A genetic network contains information about the pathway to which a gene belongs and which genes other genes interact with. Furthermore, these networks may explain the gene's function in terms of how it influences other genes and may indicate which genes are pathway initiators and therefore reveal potential drug targets [von Someren et al., 2002]. Also, understanding the position effect of genes can inform the design of
gene-therapy vectors [Hurst et al., 2004]. There is a push for current research to discover more about possible patterns and rules about gene location.

One reason for this is that these rules and relationships can give us insight into finding operons. Operons are a cluster of functionally related genes regulated and transcribed as a unit. If we can discover new clusters of coregulation, we can learn more. Genes that function at different steps of the same pathway often have the same phylogenetic profile. Phylogenetic profiling is a method of assigning functional clues to proteins based on their patterns of inheritance across multiple proteomes, and is independent of their amino acid sequence similarity to proteins of known function. These profiles may have certain predictive power in that groups of genes with the same profile are more likely to be functionally connected than those with different profiles. Also, phylogenetic profile analysis facilitates the study of protein function by assigning functional clues to uncharacterized proteins. It also has the potential to assign novel functional clues to proteins of known functions [Galperin et al., 2000].

An example of how this information can be used is in comparative genomic studies. Comparative genomic studies can reveal a high degree of conservation in genome structure. This concept is known as synteny. Synteny refers to loci contained within the same chromosome. Conserved synteny (colinearity) refers to some degree of conservation of gene content, order and orientation between chromosomes of different species or between nonhomologous chromosomes of a single species [Eckardt, 2001]. This has been observed in the gene content among the grasses, including the grain crops rice, wheat, maize, barley, sorghum, and millet. They appear to have high degrees of colinearity [Eckardt, 2001].

Also, the knowledge that is gained from my research can have an impact on other fields. The evidence obtained in the Lee and Sonnhammer (2003) study indicates that clustering of genes may not be limited to adjacent genes. This has an impact on gene order conservation studies.

The Lee and Sonnhammer (2003) results show that the area around a gene of interest should be expanded to include larger areas than the immediate neighborhood. Gene order conservation studies are important because establishing conserved gene order could be helpful in the prediction of both physical interactions between proteins and protein function [Dandekar et al., 1998]. Dandekar et al. (1998) noted that:

- If products of both genes have been tentatively assigned functions, a conserved gene order can be used to predict both physical interaction and function.
- If the function of the product of one gene in a conserved gene pair is known, it can hint at the function of the product of the neighboring gene (speed up functional assignment of unknown).
- If the functions of the products of both genes are known, the fact that the genes exist as a conserved gene pair might reveal novel functional aspects.

Conserved order of genes implies an underlying selective bias and perhaps related function. If there is positive selection for the proximity of two genes, the proximity is expected to be conserved in their orthologs. Searching for conserved gene proximity across genomes is an approach for detecting functionally coupled neighbors [Yanai et al., 2002].

### 1.2 Other Approaches

Most in research dealing with genetic networks has focused on the analysis of microarray data. There have been a multitude of modeling approaches that have been proposed for discovering genetic networks based on the analysis of the results of microarray experiments. Clustering methods are the most commonly used [Quackenbush, 2001]. These methods try to identify genes that behave in a similar manner (e.g. up regulated or down regulated) across a range of conditions or samples. Other techniques try to reconstruct and model gene-expression networks
by using models. Some of these models include but are not limited to Boolean models [D'haeseleer et al., 2000], models based on differential equations [Chen et al., 1999], or Bayesian networks [Friedman et al., 2000 and Chen, 2006]. These models are based on the notion that the expression of each gene in the network depends on the expression of some other genes, so to reconstruct such a network we have to know for each gene in the network: which genes affect it, and how they affect it.

In a typical microarray experiment, many genes in an organism are assayed under multiple conditions (i.e. different time points, different tissue samples with some common phenotype (tissue type or malignancy)). Even though microarray technology allows for studying large portions of a genome, there is only data available for a small portion of the genes, meaning a large portion of genes remain unknown. This creates a need to establish methods that facilitate the discovery of new genes and methods that help determine their functional relationships with other genes faster. Since proteins can serve different roles depending on the demands of the organism, their corresponding genes are often coexpressed with different groups of genes under different situations. This notion poses a challenge in analyzing whole genome expression data because many genes will be similarly expressed to multiple, distinct groups of genes. This means that the connections between conditionally coregulated genes are often missed because most commonly used analytical methods cannot appropriately represent these relationships. Creating a way to focus a search for coregulated genes and genetic networks that utilizes learned relationships from the genome will help with this problem. Little research has been dedicated to learning patterns the applying what is learned to focus studies.

Determining functionality of an unknown gene is commonly done using sequence information. Sequences are compared using algorithms like BLAST (Basic Local Alignment Search Tool) [Altschul et al., 1990]. The premise of these searches is that if two sequences are similar, they
probably share a common evolutionary ancestor (i.e. they are homologues) and possibly the same functionality. When there are no homologues for the given gene, in vivo experimentation through gene knockout and other assays can give researchers insight to how genes work in cells, but these studies can be labor intensive.

### 1.3 My Approach

The objective of this thesis is to develop a computational tool that uses different knowledge sources and learning algorithms to assist in focusing searches for genetic networks and help determine gene product functionality. The goal is to make the discovery of genetic networks faster and easier by limiting the search areas within the genome and to make the process of determining gene functionality faster by suggesting Gene Ontology terms for the given gene.

To build this tool, I first had to evaluate the performance of different classification learning techniques on the genomic data. The genome selected was that of Saccharomyces cerevisiae because of all the complete genomes, it has the most ideal characteristics for a test case [Lee and Sonnhammer 2003, Pál and Hurst, 2003, Cohen et al., 2000, Kruglyak, 2000 and Hurst et al., 2004].

In this research, A series of experiments was performed to determine:

1. What information is the most beneficial in classifying what networks a genomic feature belongs.
2. What information is the most beneficial in classifying genomic features to the correct Gene Ontology term.
3. If locations of interest for networks could be suggested based on a description of a feature.

Different classification learning techniques were used to analyze the genomic data. These techniques included a Bayesian classifier, a tree classifier and a lazy classifier. The Bayesian classifier used was a Naïve Bayesian classifier [Zhang, 2004]. The MultiBoostAB algorithm [Webb, 2000] using the J48 classifier [Quinlan, 1993] was the tree classifier used. The KStar classifier [Cleary and Trigg, 1995] was the lazy classifier used.

Different metrics were used to evaluate performances of these classifiers. These metrics include the percentage of correctly classified features, the Kappa statistic, precision, recall, the Fmeasure, and the area under the ROC curve. A description of all of these metrics can be found in Section 4.2.

Based on the performances of these classifiers, two rule-based systems were created. One suggests possible areas of interest where other genomic features that interact with a given genomic feature could be located. The other suggests the functionality of the given genomic feature.

To further analyze if locations of interest could be suggested based on a feature's description, a search engine was created. The search engine processed descriptions of features, retrieved features that it deemed similar, then based on the retrieved features, suggest possible networks that the feature may be involved in and areas in the genome to look for other features that may participate in these networks.

Results indicated that knowing the Gene Ontology terms of the nearest neighbors both upstream and downstream on both strands of a feature is important in determining the Gene Ontology term of that feature and determining network involvement of that feature.

### 1.4 Contributions

Through my experiments I determined, for yeast, what neighbor information was the most beneficial in classifying features to the correct network and the correct Gene Ontology term, and how many neighbors to include in this process. I found that knowing the Gene Ontology terms of a feature's two-nearest neighbors both upstream and downstream on both strands of a feature is important in determining network involvement of that feature the Gene Ontology term of that feature. Partitioning the classification process by chromosome allowed the MultiBoostAB algorithm [Webb, 2000] using the J48 classifier [Quinlan, 1993] to provide the most accurate results. The decision trees generated in the classification process became the rules in the rulebased systems I designed.

The software created acts as a filter, limiting the search area for genetic interactions for scientists and suggesting possible functionality if it is unknown. Now that I established this in yeast, this technique can be applied to other genomes to see if similar trends appear. We will then be able to determine if there are rules based on gene order or conserved synteny among the different genomes. This is beneficial to comparative genomics, and it is also helpful in understanding human disease. Animal models are commonly used in studying human disease, so discovering these relationships between animal models and humans will be easier with my software because rules can be established and then compared. Many diseases are caused by flawed networks. Being able to elucidate genetic networks provides the foundation on which new treatments or cures can be developed.

When given the location of a genomic feature, the software created can correctly suggest potential networks and thus areas of interest over $50 \%$ of the time by incorporating information about each feature's two-nearest neighbors. When only given a textual description of the feature, the software recommended the correct network and thus areas of interest about $43 \%$ of the time.

These results can help with experimental design. Little effort has been placed on analyzing the genome for network relationships before initiating the microarray experiments. My software can help design (i.e. which probe sets should be included on the microarray chip) the actual chips. The software designed here can provide researchers with a way to focus their search for coregulated genes, which will lead to better chip design, and ultimately save the researcher time and money. Scientists can then use my software to maximize the possibility that their microarray experiment represents a good coverage of potential gene interactions. This work could be pivotal for research by providing a more focused experimental approach to elucidate biological roles and relationships between genes.

When given the location of a genomic feature, the software created can correctly suggest a gene product's Gene Ontology term(s) over $80 \%$ of the time by incorporating information about each feature's two-nearest neighbors. This does not eliminate the need to confirm actual functionality in the lab, but it does help by limiting the list of possible functions of a gene product.

If we make the analogy that a genome is a collection of neighborhoods, and a genomic feature is like a house, then knowing your address and what your two-nearest neighbors on both sides of your house and across the street do tells us a lot about what you do (Gene Ontology terms) and what organizations (networks) you might be involved in. The software created in this study takes advantage of this phenomenon making it easier to come up with reliable hypotheses about the location of potential genetic regulatory networks within a genome and about possible functionality of a gene.

## 2 Background

### 2.1 Knowledge Sources

Currently, there are several biological databases that are open to the public. These databases are: Gene Ontology (GO), Saccharomyces Genome Database (SGD) and KEGG (Kyoto Encyclopedia of Genes and Genomes).

### 2.1.1 Saccharomyces Genome Database

SGD is a scientific database that contains information about the molecular biology and genetics of the yeast Saccharomyces cerevisiae. Saccharomyces cerevisiae is also known as baker's or budding yeast. This database contains the information about the known open reading frames and genes. Maps of all the sixteen chromosomes and the mitochondrial chromosome can be obtained from this database (Hong et al., 2006). Summaries of several genomic features can be accessed at SGD.

### 2.1.2 Gene Ontology

Gene Ontology is not a database of gene sequences, nor a catalog of gene products. Rather, Gene Ontology describes how gene products behave in a cellular context using three different ontologies: molecular function, biological process, and cellular component. These classifications were chosen by the Gene Ontology Consortium because they represent information about genes and gene products [Gene Ontology Consortium, 2000]. The ontology comprises a set of welldefined terms with well-defined relationships. The structure of Gene Ontology reflects the current representation of biological knowledge as well as serving as a guide for organizing new data [Gene Ontology Consortium, 2000]. The vocabulary is fluid and consistently undergoes revision. The intention of the Gene Ontology is to make possible, in a flexible and dynamic way, the annotation of homologous gene and protein sequences in multiple organisms using a
common vocabulary. Sharing this common vocabulary will result in the ability to query and retrieve genes and proteins based on their shared biology [Gene Ontology Consortium, 2000].

### 2.1.3 KEGG

KEGG contains knowledge about known biological (regulatory, metabolic and cellular) pathways, thus it contains knowledge about genetic networks. KEGG can be indexed based on gene products. Information on the pathways in which the gene/gene product participates, the other genes/gene products in those pathways, and locations for all the genes involved can be determined.

### 2.2 Learning

Learning encompasses a broad range of processes. There are many definitions of learning. In general, it can be stated that learning occurs when an agent changes its structure, program, or data (based on inputs or in response to external information) in such a manner that its expected future performance improves [Nilsson, 1996]. Dietterich (1986) states that learning can be defined as an increase in knowledge. Knowledge is then defined by Dietterich as the following: An agent (program) knows a fact $F$ if the agent has been told $F$ or if the agent can logically infer $F$ from its other knowledge. This form of knowledge can be called "knowledge in principle" or "deductive closure knowledge". The logical inferences are assumed to preserve correctness (i.e. they are monotonic, deductive inferences) [Dietterich, 1990]. Learning can occur under two circumstances with this definition of knowledge.

1. When an agent is told a fact $F$ that it did not know
2. When an agent makes an "inductive leap" and chooses to believe some fact $F$ that is not entailed by its existing knowledge [Dietterich, 1990].

The type of learning that is of interest for this research in induction. Induction uses specific examples to reach general conclusions. Inductive learning methods can be defined as those methods that systematically produce general descriptions or knowledge from the specific knowledge provided by domain examples. This type of learning will assist in determining if there are patterns in gene location and if there are rules that can be found in genetic network data. Specifically, data mining methods will be used.

Data mining is defined as the process of discovering patterns in data. Patterns allow us to make nontrivial predictions on new data. It is helpful when the patterns that are mined are represented in terms of a structure that can be examined, reasoned about, and used to inform future decisions. These are structural patterns because they capture the decision structure in an explicit way, thus, they help to explain something about the data. There are different styles of learning in data mining applications. These styles include classification learning, association learning, clustering and numeric prediction.

### 2.2.1 Classification learning

Classification learning is the learning style used in this research. Based on the description provided by Witten and Frank (2005), in classification learning, the learning system is presented with a set of classified examples. From these examples, the system is expected to learn a way of classifying unseen examples.

There are several data mining algorithms that perform classification learning. These can be broken down into various categories. Some of the more commonly known categories are: Bayesian classifiers, tree classifiers and functions (classifiers that can be written down as mathematical equations).

### 2.2.1.1 Bayesian classifiers

Bayesian classification is based on Bayes' Theorem. Bayesian classifiers are useful in predicting the probability that a sample belongs to a particular class or grouping. A common of these types of classifiers are called Naïve Bayes classifiers. A more detailed discussion follows in Section 2.2.2.1.

### 2.2.1.2 Tree classifiers

Decision tree classifiers are hierarchically based classifiers that compare the data with a range of properly selected features. The selection of features is determined from an assessment that depends on the algorithm being implemented. Assessment can be based on the spectral distributions or separability of the classes in the dataset. Decision trees are often referred to as a "divide-and-conquer" method. Nodes in a tree pertain to testing a particular attribute. Leaf nodes give the classification that can be applied to all instances that reach the leaf, or a set of classifications, or a probability distribution over all possible classifications. In the classification process of an unknown instance, the instance is routed down the tree according to the values of the attributes tested in successive nodes, and when a leaf is reached, the instance is classified based on the class assigned to that leaf [Witten and Frank, 2005]. Basically, a series of questions separate data into increasingly narrower categories. The goal is to ask the proper sequence of questions that will accurately model patterns in the data. The advantages of the decision tree classifier are that computing time is less than a maximum likelihood classifier and by comparison the statistical errors are avoided. However the disadvantage is that the accuracy depends fully on the design of the decision tree and the selected features/attributes.

### 2.2.1.3 Functions

These types of classifiers are usually based on logistic regression. Logistic regression attempts to produce accurate probability estimates by maximizing the probability of the training data. Accurate probability estimates lead to accurate classifications [Witten and Frank, 2005]. Any regression technique (linear or non-linear) can be used for classification. To do this, regression must be performed for each class, setting the output equal to one for training instances that belong to the class and zero for those that do not. The result is a linear expression for the class. To classify an unknown, calculate the value of each linear expression and choose the one that is largest. Other approaches learn a hyperplane that separates the instances pertaining to the different classes.

### 2.2.1.4 Lazy Classifiers

The classifiers that are mentioned above are sometimes considered "eager" because they precompute (perform statistics, build trees, etc) over the entire dataset. They construct a generalization model before receiving any new samples to classify. This makes the training of the classifier a timely activity, but prediction of the classification of a new instance is fast. Lazy classifiers do no precomputation, and they only focus on a particular set of the data. There is no training, thus the prediction of the classification of a new instance is slow because classifiers are not built until new samples need to be classified. Lazy classification uses a richer hypothesis space, which can improve classification accuracy, but the attributes are all equally weighted, which can decrease classification accuracy. A major disadvantage of lazy classification is that all training samples need to be stored, which leads to expensive storage costs and requires efficient indexing techniques. The most common of these types of classifiers use a distance function to classify instances.

### 2.2.2 Classification Algorithms

The types of algorithms described above for classification learning can be found in WEKA (the Waikato Environment for Knowledge Analysis). WEKA is open source software issued under the GNU General Public License. It is a collection of machine learning algorithms for data mining tasks. WEKA contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization [Witten and Frank, 2005].

The specific algorithms that this research is interested in are the Bayesian classifier technique called Naïve Bayes, the tree classifier algorithm known as C4.5, and the lazy classifier known as KStar. In WEKA, C4.5 is referred to as J48. To prevent the over-fitting of data, the MultiBoostAB algorithm using the J48 classifier was used.

### 2.2.2.1 Naïve Bayes

A Naïve Bayes classifier is based on models that incorporate strong independence assumptions. Naïve Bayes classification does not require that you have lots of observations for each possible combination of the attributes. Rather, the attributes are assumed to be independent of one another. In other words, Naïve Bayes classifiers assume that the effect of an attribute value on a given class is independent of the values of other attributes. This assumption is called class conditional independence. It is made to simplify the computation and in this sense considered to be "Naïve".

These assumptions make for very efficient algorithm for classification. Naïve Bayesian classifiers generate rules based on Bayes' rule of conditional probability. It uses all the attributes and allows them to make contributions to the classification as if they were all equally important and independent of one another.

To explain how the Naïve Bayesian classifier, we need to consider a Bayesian classifier. Let us say an instance, $I$, in classification is represented by a tuple of attribute values $\left(x_{1}, x_{2}, \ldots, x_{n}\right)$, where $x_{i}$ is the value of attribute $X_{i}$. Let us assume that there are only two classes, positive or negative. Next, we let $C$ represent the classification variable, and let $c$ be a value of $C$. According to Bayes rule, the probability of an instance $I=\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ being in class $c$ can be calculated:

$$
p(c \mid I)=\frac{p(I \mid c) p(c)}{p(I)}
$$

The instance, $I$, is classified as the class positive $C$ if and only if:

$$
f_{b}(I)=\frac{p(C=\text { positive } \mid I)}{p(C=\text { negative } \mid I)} \geq 1
$$

The function, $f_{b}(I)$, is considered a Bayesian classifier (Zhang, 2004).

When the assumption is made that all the attributes are independent given the value of the class variable (conditional independence), we get:

$$
p(I \mid c)=p\left(x_{1}, x_{2}, \ldots, x_{n} \mid c\right)=\prod_{i=1}^{n} p\left(x_{i} \mid c\right)
$$

This results in a Naïve Bayesian classifier function:

$$
f_{n b}(I)=\frac{p(C=\text { positive })}{p(C=\text { negative })} \prod_{i=1}^{n} \frac{p\left(x_{i} \mid C=\text { positive }\right)}{p\left(x_{i} \mid C=\text { negative }\right)}
$$

(Zhang, 2004).

The Naïve Bayes classifier has several properties that make it surprisingly useful in practice, despite the fact that the far-reaching independence assumptions are often violated. Like all probabilistic classifiers under the MAP decision rule (maximum a posteriori decision rule: a rule that picks the hypothesis that is most probable), it arrives at the correct classification as long as the correct class is more probable than any other class. The class probabilities do not have to be estimated very well because the overall classifier is robust enough to ignore serious deficiencies in its underlying naïve probability model. These types of classifiers, despite their over-simplified assumptions, often work much better in many complex "real-world" situations. Studies comparing classification algorithms have found the Naïve Bayesian classifier to be comparable in performance with classification trees and with neural network classifiers [Stewart, 2002]. They have also exhibited high accuracy and speed when applied to large databases [Zhang, 2004 and Witten and Frank, 2005].

Since attributes in these classifiers are treated as if they were independent, the addition of redundant attributes skews the learning process. Also, dependencies between attributes reduce the ability of the Naïve Bayes classifier to discern what is actually going on.

### 2.2.2.2 MultiBoostAB using the J 48 classifier

MultiBoostAB is the WEKA version of the MultiBoosting algorithm. MultiBoosting combines AdaBoost (a boosting technique) [Freund and Schapire, 1996] with wagging (a variant of bagging) [Bauer and R. Kohavi, 1998] to form decision committees. This technique "boosts" a learning algorithm to a stronger learning algorithm by taking advantage of AdaBoost's high bias and variance reduction and wagging's strong variance reduction. It has been shown that when using the C4.5 algorithm as its base learning algorithm, MultiBoosting produces decision committees with a lower error than either AdaBoost or wagging [Webb, 2000].

This technique allows us to combine multiple models/classifiers generated by the learning algorithm. Bagging (Bootstrap aggregating) and boosting create an ensemble of classifiers from a dataset. Each of the classifiers is built using a different training set obtained from the original dataset using resampling techniques. Combining the decisions of different classifiers into one prediction is done by taking a vote in both bagging and boosting. They differ in how they derive the individual models. In bagging, an equal weight is assigned to the models. In boosting, weighting is used to give more influence to the more successful models.

Bagging works by producing replications of the training set by sampling with replacement. Each of these replications has the same size as the original set, but some instances can appear more than once and some do not appear at all. A classifier (like a decision tree) is generated for each replication, and they are used on each instance of the test set. For final classification, each of the classifiers votes for each class [Witten and Frank, 2005].

Wagging differs from bagging in that it assigns random weights to the instances in the training set. In the MultiBoosting algorithm, random instance weights are assigned using the continuous Poisson distribution. The weights are calculated using the formula:

$$
\text { Poisson( })=-\log \left(\frac{\text { Random }(1 \ldots 999)}{1000}\right)
$$

where Random(min.max) returns a random integer value between min and max inclusive [Webb, 2000].

Boosting differs from bagging in that the process is iterative. In bagging, the models/classifiers are built separately. In boosting, each new classifier is influenced by the performance of those built before it. This encourages new classifiers to become "experts" for instances handled incorrectly by the earlier classifiers.

The variant of boosting that MultiBoosting incorporates is AdaBoost [Freund and Schapire, 1996]. It works by first assigning an equal weight to all training instances. For each iteration, it calls the learning algorithm to form a classifier for the weighted dataset and it stores the resulting model. The error $e$ of the model is computed. If e is equal to zero, or e is greater than or equal to 0.5 , model generation is terminated. Next, each instance in the dataset gets reweighted according to the classifier's output. The weights of the correctly classified instances get decreased by using the following formula:

$$
\text { weight }=\text { weight } *\left(\frac{e}{(1-e)}\right)
$$

The weights of the incorrectly classified instances remain unchanged. This produces a set of "easy" instances with low weight and a set of "hard" instances with high weight.

In the subsequent iterations, a classifier is built on the reweighted dataset which focuses on classifying the "hard" instances correctly. To form a prediction, the output of the classifiers from each iteration is combined using a weighted vote. Classifiers that perform well on the weighted training data (e close to zero) receive a higher weight, and classifiers that perform poorly (e close to 0.5 ) receive a lower weight. The weight of the voting is determined by the formula:

$$
\text { weight }=-\log \left(\frac{e}{(1-e)}\right)
$$

To make the prediction, the weights of all the classifiers that vote for a particular class are summed, and the class with the greatest sum is chosen [Frank and Witten, 2005].

Boosting often produces classifiers that are significantly more accurate than bagging, but due to overfitting the data, boosting can generate classifiers that are significantly less accurate than a
single classifier built from the same training set [Frank and Witten, 2005]. This motivated the creation of MultiBoosting.

MultiBoosting should be used if the learning algorithm is unstable. When decision trees are built for a dataset, the process is unstable, meaning slight changes in the training data may result in different attributes being chosen at a particular node. This has an impact on the structure of the subtree beneath that node. On algorithms that are stable, like k-nearest neighbors, this technique should not be applied because it degrades the estimates of the algorithm.

J48 is the WEKA version of C4.5 [Quinlan, 1993]. C4.5 is a decision tree generating algorithm based on the ID3 algorithm [Quinlan, 1986]. C4.5 is an extension of ID3 that accounts for unavailable values, continuous attribute value ranges, pruning of decision trees and rule derivation. In building a decision tree, C4.5 can deal with training sets that have records with unknown attribute values by evaluating the gain, or the gain ratio, for an attribute by considering only the records where that attribute is defined. In using a decision tree, it can classify records that have unknown attribute values by estimating the probability of the various possible results [Witten and Frank, 2005].

In building decision trees, the algorithm needs to determine which attribute to split on given a set of examples with different classes. This attribute is selected based on information [Shannon and Weaver, 1949]. Information is measured in bits. When information is associated with a node, it represents the expected amount of information that would be needed to specify how a new instance should be classified given that instance reached that node. It is calculated using the formula:

$$
I=\left(P\left(v_{1}\right), \ldots, P\left(v_{n}\right)\right)=\sum_{i=1}^{n}-P\left(v_{i}\right) \log _{2} P\left(v_{i}\right)
$$

To simplify, suppose the training set contains positive ( $p$ ) and negative ( $n$ ) examples, then the estimate of the information contained in the correct classification is:

$$
I\left(\frac{p}{p+n}, \frac{n}{p+n}\right)=-\frac{p}{p+n} \log _{2} \frac{p}{p+n}-\frac{n}{p+n} \log _{2} \frac{n}{p+n}
$$

Calculating how much information that is still needed after the attribute test needs to be determined. Any attribute divides the training set into subsets according their values for that attribute (an attribute can have $v$ distinct values). Each subset has positive ( $p_{i}$ ) and negative ( $n_{i}$ ) examples. So on average, after testing attribute $A$, we will need:

$$
\operatorname{Remainder}(A)=\sum_{i=1}^{v} \frac{p_{i}+n_{i}}{p+n} I\left(\frac{p_{i}}{p_{i}+n_{i}}, \frac{n_{i}}{p_{i}+n_{i}}\right)
$$

bits of information to classify the example. The information gain from the attribute test is then calculated as:

$$
\operatorname{Gain}(A)=I\left(\frac{p}{p+n}, \frac{n}{p+n}\right)-\operatorname{Remainder}(A)
$$

The equations for these calculation are taken from Russell and Norvig (2003).

C4.5 builds a tree by selecting the best attribute to split the dataset on. For each attribute, it calculates the information gain from splitting the tree on that attribute. The best attribute has the highest information gain. A decision node is created that splits the dataset on that best attribute. The process is repeated on the subtrees of that node.

### 2.2.2.3 KStar

KStar is an instance-based classifier. An instance-based classifier classifies a test instance by finding the training instance(s) that is/are the most similar to the test instance. The KStar algorithm is a member of the $k$-nearest neighbor classifiers. $K$-nearest neighbor classifiers classifies each instance by looking at the nearest $k_{k}$ data points and determining the class an instance belongs to based on the most common class amongst the nearest $k$ data points. The nearest neighbors of an instance are defined in terms of a standard distance. KStar uses an entropy-based distance function. This approach is motivated by information theory [Cleary and Trigg, 1995].

The distance between two instances is defined by the complexity of transforming one instance into another. The KStar distance sums over all the possible transformations between two instances. Instances are classified into certain categories based on probability. The probability that an instance, $i$, belongs to a category $C$ is calculated by summing over the probabilities from $i$ to each instance that is a member of C :

$$
\operatorname{PStar}(C \mid i)=\sum_{b \in C} \operatorname{PStar}(b \mid i)
$$

The probabilities for each category are calculated. The category with the highest probability is selected as the classification of the new instance [Cleary and Trigg, 1995].

The KStar requires the user to specify a blend factor. It specifies how the distance function acts by setting the guideline for $k$ (how many instances to consider when classifying). A lower blend factor will classify an instance based on a more specialized set of training data. A higher blend factor will classify an instance based on a more general set of training data [Cleary and Trigg, 1995].

The user also needs to specify how to handle missing values. There are four options. The missing values can be ignored when computing the distance. The missing values can be interpreted as the maximum possible distance between the instances of a particular attribute. Before the distance is computed, the missing values can be normalized over the attribute. The distance for the missing values can be determined from the average column entropy curve [Cleary and Trigg, 1995].

KStar is a simple method that is generally very accurate. It is slow because the entire set must be searched so that the k-nearest neighbors can be found. What may skew results is the fact that KStar treats all attributes as equally important [Cleary and Trigg, 1995].

### 2.3 Information Retrieval

In this research, similar genomic functions were determined using Information Retrieval (IR) techniques using the vector-space model (VSM). The VSM is a common approach in IR for similar-document retrieval [Salton, 1971]. In the VSM, each document is represented by a list (vector) of weighted terms. For example document $i$ is equal to:

$$
D_{i}=\left(w_{i 1}, w_{i 2}, \ldots, w_{i n}\right)
$$

where $w_{i n}$ is the weight assessed to term $n$ in the document $D_{i}$.

When analyzing a term in a document, how well the term describes its document, and the term's scarcity need to be included. These two concepts are involved in determining a term's weight. This weight is associated with and describes a term's value for a document. The weighting system for each term in the document uses a $t f$-idf scheme. ( $t f=$ term frequency; $i d f=$ inverse document frequency). In this term-weighting scheme, the $t f$ and $i d f$ are calculated in the following manner:

$$
\begin{gathered}
t f=\frac{\text { frequency of the term in the document }}{\text { frequency of the most frequent word in the document }} \\
\text { idf }=\log _{10}\left(\frac{\text { total number of documents in the collection }}{\text { the number of documents in the collection that contain the term }}\right)
\end{gathered}
$$

Thus the weight of a term is calculated by:

$$
\text { Weight }=t f * i d f
$$

A similarity measure can be computed between pairs of weighted term vectors that reflects text similarity. Using the VSM makes it possible to compare two documents using vector algebra, as, for example, the cosine measure of similarity [Salton and McGill, 1983]. In this model, the similarity of document vector, d 1 , to another document vector, d 2 , equals the cosine of the angle between them.


$$
\operatorname{similarity}(d 1, d 2)=\operatorname{cosine} \theta=\frac{d 1 \cdot d 2}{|d 1||d 2|}
$$

Cosine is the normalized dot product. The degree of similarity is determined by the cosine of the angle; the smaller the angle, the more similar the documents.

### 2.4 Patterns in Genomes

Several research groups have noted that genes with related function are often located close to each other on the chromosomes [Jensen et al., 2003]. The existence of operons makes this grouping trend stronger in prokaryotes. Operons are a common form of gene organization in
bacteria and archea. In bacterial operons, genes are coexpressed from a single promoter to make a polycistronic mRNA (an mRNA that codes for more than one protein).

An example of an operon is the lac operon. The lac operon was the first discovered example of tight integration of gene expression and function. The lac operon is a transcriptional unit of genes functioning in a common metabolic pathway in bacteria utilizing lactose. The genes encode a hydrolase cleaving lactose, lacZ, a galactoside permease, lac Y, and a galactoside transacetylase, lacA. The coexpression of these genes is achieved by their organization in a linear transcription unit that is regulated by a repressor.

Many times the genes contained within an operon are functionally related, thereby facilitating coregulation [Blumenthal, 1998]. It was believed that eukaryotic genes appeared to be transcribed without apparent organization by function or by temporospatial expression. This view is being slightly altered by the presence of examples of clustering of genes of related function (e.g. Hox genes, hemoglobin genes and immunoglobulin genes) and by the presence of tight clustering of genes whose functions are not obviously related [Blumenthal, 1998]. It has also been observed that an interesting feature of the $A$. thaliana genome is the frequent occurrence of clusters of closely related protein-coding genes adjacent on the same strand [Wambutt et al., 2000]. The given examples hint that the eukaryotic organization of genes on chromosomes may not be as random as once believed.

Blumenthal (1998) stated that there are operons in eukaryotes that are clear examples of coregulation of genes whose products need to be coexpressed temporarily and spatially. It is important to note that the observed operons in eukaryotes are different from those observed in prokaryotes in terms of mechanism and gene content [Blumenthal, 1998]. Operons that have been detected in eukaryotes differ from those in prokaryotes in that eukaryotic polycistronic mRNAs are not translatable directly as in prokaryotes [Kozak, 1999].

Lee and Sonnhammer (2003) noted that there could exist some degree of selection for keeping coregulated genes in the same region of a chromosome. For instance, the gene positioning can make the genes (as a group) more efficiently available to transcription. This is supported by the fact that coregulated human genes are often linked functionally in synexpression groups [Niehrs and Pollet 1999]. Synexpression groups designate sets of genes that share a complex 'spatial' expression pattern (multiple tissues), and that function in the same process.

Képès (2003) found that in yeast, genes that are regulated by the same sequence-specific transcription factor tend to be regularly spaced along chromosomes. Co-expressed gene pairs have also been found to be regularly spaced along chromosomes in Saccharomyces cerevisiae [Mannila et al., 2002] and in Drosophila melanogaster [Balázsi et al. 2003].

Positional clustering of coexpressed genes has been studied in Caenorbabditis elegans. After exclusion of operon and tandem duplication (the major mechanisms for observed coexpression), gene clustering is still evident [Roy et al., 2002 and Lercher et al., 2003].

Cho et al. (1998) first showed clustering of co-expressed genes in yeast on a genomic scale when they found that $25 \%$ of genes with cell-cycle dependent expression patterns were directly adjacent to genes induced in the same phase of the cell cycle. Several genes in yeast that are in co-expression clusters seem to be functionally related (i.e. they belong to the same Gene Ontology classification [Fukuoka et al., 2004] or the same MIPS (Munich Information Centre for Protein Sequences) category) [Cohen, et al., 2000].

Caron et al. (2001) found highly expressed genes co-localized in Homo sapiens. Caron et al. (2001) presented work on the Human Transcriptome Map. This map allows us to view the whole chromosome. When inspected, whole chromosome views reveal a high order organization of the human genome. In humans, it has been suggested that expression neighborhoods serve to
regulate housekeeping functions [Lercher et al., 2002]. From the Transcriptome Map, we can see most of the chromosomes show clusters of highly expressed genes, which are referred to as RIDGEs (regions of increased gene expression) (see the Figure 2.1 below from the paper by Caron et al. (2001)). Since there is a continuum from small to very large clusters, it is hard to quantitatively define a RIDGE. This study shows that RIDGEs most likely represent a higher order of structure in the genome.


Figure 2.1. From Caron et al. Regional expression profiles for 23 human chromosomes show a clustering of highly expressed genes in RIDGEs. Expression levels are shown as a moving median with a window size of 39 genes. There are 74 regions with one or more consecutive moving medians that have a lower limit of four times the genomic median; 27 of them have a length of at least 10 consecutive moving medians (indicated by the bars to the right).

A study by Gabrielsson et al. examined genes expressed in human adipose tissue via microarray analysis. They observed clusters of adipose specific genes when they mapped these genes back
to the human genome [Gabrielsson et al., 2000]. This study suggests that clusters of tissuespecific genes do exist. This type of phenomenon was also demonstrated within the mouse [Li et al., 2005].

Spellman and Rubin (2002) describe a transcriptional profiling study that reveals a surprising correlation between the organization of genes along Drosophila chromosomes and their expression levels. This study showed that more than $20 \%$ of genes clustered into co-regulated groups of 10-30 genes in the Drosophila genome. They noted that neighborhoods composed of an average of 15 contiguous genes show markedly similar relative expression levels. Although the average neighborhood contains 15 genes, there is a very wide range. These neighborhoods are not obviously composed of genes with related functions that might be expected to exhibit coregulation, as is the case for the rRNA, histone, Hox, and globin gene clusters [Spellman and Rubin, 2002].

With these findings, Lee and Sonnhammer (2003) examined gene clustering of pathways in eukaryotes. The research presented in the Lee and Sonnhammer (2003) study is based on genes in metabolic pathways as defined in the Kyoto Encyclopedia for Genes and Genomes (KEGG), with missing enzymes filled in by homology. The species included in this study were: Homo sapiens, Caenorbabditis elegans, Drosophila melanogaster, Arabidopsis thaliana, and Saccharomyces cerevisiae. This study analyzed which pathways are significantly clustered in each species, and whether the clustered pathways are the same in different species. The clustering level was measured by calculating the overall degree of gene colocalization compared to random. The 'clusters' presented in this paper are thus not necessarily compact. The clusters may correspond to rather large regions with high concentrations of pathway members, although nonmembers may also be present in such regions [Lee and Sonnhammer, 2003].

To determine whether genes in a pathway are found in closer proximity than expected by chance, Lee and Sonnhammer (2003) developed a formula for calculating a clustering score. They noted that such a formula needs to fulfill a set of criteria:

1. The score should increase monotonously with increased proximity
2. The score between genes on different chromosomes needs to be defined and to be the same as the average score between two randomly placed genes on an average chromosome.
3. The score should be normalized for genome and pathway size to allow comparative studies.

To estimate the significance of the observed average clustering score of a pathway in a genome, it was compared to the distribution of 200 iterations of placing the same number of genes on randomly picked known gene positions. Friedman and Hughes (2001) used a similar approach for determining whether observed gene patterns are expected by chance [Lee and Sonnhammer, 2003].

The results of the Lee and Sonnhammer (2003) work show between $98 \%$ and $30 \%$ of the analyzed pathways in a genome were found to exhibit significantly higher clustering levels than expected by chance. The Saccharomyces cerevisiae genome showed the most clustering followed by Homo sapiens, Caenorbabditis elegans, Arabidopsis thaliana, and Drosophila melanogaster in descending order. Surprisingly, Lee and Sonnhammer (2003) did not observe much agreement between genomes in terms of which pathways are most clustered. Only seven of 69 pathways found in all species were significantly clustered in all five of them [Lee and Sonnhammer, 2003].

In their study, Lee and Sonnhammer (2003) frequently found that genes in a pathway are in closer proximity than would be expected by chance. They also noted that genes that are 'clustered' are much more distant than those detected in prokaryotes; however there may still be
an advantage for functionally related eukaryotic genes to be close in a genome even if separated by other genes [Lee and Sonnhammer, 2003]. From this study, it seems that the clustering of pathway members must be a selected trait. In further support of this, there is no detectable correlation between sequence similarity and clustering score in the Lee and Sonnhammer work [Lee and Sonnhammer, 2003].

The notion that genes of comparable and/or coordinated expression cluster is important in terms of understanding how genomes function and how the genomes have evolved.

### 2.5 Possible explanation of clusters/groups of genes in genomes

The mechanism(s) responsible for the observed similarities in expression of adjacent genes was not revealed in the Spellman and Rubin (2002) study. The authors stated that the results obtained were most consistent with regulation at the level of chromatin structure based on the following reasons:

1. The regions showing similarities in expression are quite large (average 15 genes) with each presumably having its own core promoter.
2. It is frequently the case that one or two genes in a group display a high level of differential expression. They state that "if the chromatin in a region of the chromosome that contained many genes was 'opened' so that a single target gene could be expressed, it might increase the accessibility of promoters and enhancers of other genes to the transcriptional machinery, leading to a modest parallel increase in their expression." [Spellman and Rubin, 2002]

Oliver et al. (2002) find the Drosopbila data presented by Spellman and Rubin (2002) warrant closer examination to try to discover what could underlie the observed similarity of gene expression within neighborhoods.

Oliver et al. (2002) explored explanations for these neighborhoods. The simplest explanation is that co-regulation within an expression neighborhood may be due to incidental interactions between promoters and transcriptional enhancers (see Figure 3). In this model, transcription of one or more genes in a genomic cluster is regulated by the usual suspects (transcription factors) binding at the appropriate sites and activating nearby genes as well as the target gene. The resulting inappropriate expression of genes other than the target is tolerated because it has little biological effect. Transcription factors have a limited range of effect [Dorsett, 1999]. This means if strong activators are responsible then one might expect to see a steep fall-off in the effects of a given factor with distance from its core binding site (See the curve in Figure 2.2 below). But Oliver et al. (2002) state that the data presented by Spellman and Rubin (2002) suggest that in fact the pattern of gene expression within a neighborhood is essentially a 'square wave' (See the curve in the Figure 2.3 below).


Figure 2.2. Incidental regulation. A transcription factor (green oval) binds at a target gene (green arrow) and incidentally up-regulates neighboring genes. In this model, the level of expression of neighboring genes is determined by proximity to the target gene and is expected to decrease with distance from the target gene (the green line at the top of each panel indicates the gene expression profile across the neighborhood) [Oliver et al., 2002].


Figure 2.3. A structural domain model. A discrete 'open' chromatin domain is created as a result of activation of a target gene within the domain. Flanking boundary or insulator elements (yellow ovals) define the neighborhood and the limits of the open chromatin domain (Note the 'square wave' expression profile) [Oliver et al., 2002].

In the Spellman and Rubin (2002) paper, they favored a structural chromatin domain model (See Figure 2.3 above), involving the opening of the chromatin of an entire neighborhood as a result of activation of a target gene within the neighborhood. Spellman and Rubin (2002) stated that the creation of a domain of open chromatin structure would increase the availability of the promoters and enhancers of all the genes in the neighborhood to the transcriptional machinery. This would then lead to correlated increases in expression. Such a domain of open chromatin structure could be delimited by boundary elements or insulators, accounting for the square wave profile (See Figure 2.3 above) [Oliver et al., 2002].

Oliver et al. (2002) find a problem with this model. They feel that the increased chromatin accessibility is just as likely to facilitate the binding of repressors as activators, the result is that some genes would be up-regulated and some down-regulated. This is not consistent with neighborhoods of co-regulation. But if increased accessibility primarily affects basal (non-
activated) expression, there could be a general increase in transcription of all the genes in the neighborhood [Oliver et al., 2002].

The eukaryotic genome requires extensive compaction of DNA to fit into the nucleus. Packing is accomplished by wrapping the DNA around histones and other proteins. When genes are transcribed, a great deal of unpacking of regions of DNA is required.

The region of DNA which is unfolded contains a number of genes. It has been suggested that keeping functionally related genes near, even if not adjacent, may ease the burden of the unpacking of DNA in the cell for transcription [Lee and Sonnhammer, 2003].

### 2.6 Current Approach

### 2.6.1 Determining Networks

Thus far, a multitude of modeling approaches has been proposed for discovering genetic networks. These efforts focus on the analysis of the results of microarray experiments. Clustering methods are the most commonly used [Quackenbush, 2001]. These methods try to identify genes that behave in a similar manner (e.g. up regulated or down regulated) across a range of conditions or samples. There are various clustering techniques that have been applied to gene expression data. The most popular technique is hierarchical clustering [Eisen, 1998]. Other techniques include $k$-means clustering [Bittner et al., 2000 and Datta and Datta, 2003], fuzzy $k$ means clustering [Gasch and Eisen, 2003], self-organizing maps [Tamayo et al., 1999], deterministic annealing [Alon et al., 1999], and graph theoretical approaches [Sharan and Shamir, 2000]. The quality of results from clustering algorithms is highly dependent on the input (i.e., on the data that is used for the analysis). It is not uncommon that more than one clustering technique is applied to a dataset. Often the results are not considered valid unless the clusters can be reproduced by various techniques. This gives confidence that the patterns observed are
true biological phenomena that are independent of the analysis technique [Kaminski and Friedman, 2002].

There are several hierarchical clustering algorithms. They differ in the way they calculate distance. Michael Eisen (1998) has developed one of the most popular hierarchical clustering and visualization tools (Cluster and Treeview). This algorithm is an agglomerative process in which single member clusters are fused to bigger and bigger clusters based on a similarity metric. This algorithm begins by computing a pairwise distance matrix between all the genes, the distance matrix is then explored for the nearest genes, and these genes are defined as a cluster. After a new cluster is formed by agglomeration of two clusters, the distance matrix is updated to reflect its distance from all other clusters. The algorithm then searches for the nearest pair of clusters to agglomerate, and the process repeats. The object of this algorithm is to compute a dendrogram, in which multiple clusters are fused in nodes according to their similarity, that resulting in a single hierarchical tree [Eisen et al., 1998 and Kaminski and Friedman, 2002].
$k$-means clustering is an iterative process. This algorithm searches for clusters that are defined in terms of their "center" points or means. The number of clusters that are going to be generated needs to be fixed in advance for this algorithm. Underestimating $k$ will result in large clusters of many genes that display divergent gene-expression patterns, while overestimating $k$ will over-fit the data and split groups of similarly expressed genes into multiple, small clusters. Usually another clustering algorithm, such as hierarchical clustering, is run to determine the initial cluster centers to be used in this algorithm. This method does not generate a hierarchical tree, but rather a predetermined number of clusters, so $k$-means clustering is especially useful in cases in which one knows how many distinct gene expression patterns to expect [Datta and Datta, 2003]. Once a set of cluster centers is defined, each gene is then assigned to the cluster it is closest to. The clustering algorithm then adjusts the center of each cluster of genes to minimize the sum of
distances of genes in each cluster to the center. This results in a new choice of cluster centers, and so we can reassign genes to clusters and repeat the process. These iterations are applied until convergence [Kaminski and Friedman, 2002].

Fuzzy $k$-means clustering algorithms have also been used to identify clusters of genes. The essential difference between fuzzy $k$-means clustering and traditional $k$-means clustering is the partitioning of genes into clusters. Rather than the hard partitioning used in the traditional $k$ means clustering, where genes belong to only a single cluster, fuzzy $k$-means clustering considers each gene to be a member of every cluster, with a variable degree of 'membership' [Gasch and Eisen, 2003]. This provides fuzzy $k$-means clustering with the benefit of being less sensitive to over-fitting, because the genes are not forced to belong to only a single cluster. In this algorithm, each gene has a total membership of 1.0 that is apportioned to clusters on the basis of the similarity between the gene's expression pattern and that of each cluster centroid. Genes whose expression patterns are very similar to a given centroid will be assigned a high membership in that cluster, whereas genes that bear little similarity to the centroid will have a low membership [Gasch and Eisen, 2003]. Using this algorithm, genes can be assigned significant memberships to more than one cluster. This can help reveal genes whose expression is similar to multiple, distinct groups of genes.

Tamayo et al. (1999) introduced the use of self-organizing maps (SOMs) in the analysis of microarray data. SOMs are preferable when dealing with "messy" data that contains outliers and irrelevant parameters. Similar to the $k$-means procedure, the genes are assigned into a predetermined set of clusters. However, unlike $k$-means, what follows is an iterative process in which gene expression vectors in each cluster are "trained" to find the best distinctions between the different clusters. In other words, a partial structure is imposed on the data and then this structure is iteratively modified according to the data [Kaminski and Friedman, 2002].

In deterministic annealing, genes are input vectors that are assigned to clusters based on what Rose (1998) called the association probability in the probabilistic model he defined in his paper. This technique has some similarity to fuzzy clustering, where each gene has partial membership in clusters. However, the cluster formulation is purely probabilistic. Clusters are considered as regular (nonfuzzy) sets whose exact membership is the outcome of a random experiment. Alon et al. (1999) used deterministic annealing to perform hierarchical clustering. It was used to organize gene data in a binary tree.

Graph-theoretic clustering is an innovative clustering method that utilizes graph-theoretic and statistical techniques to identify tight groups of highly similar elements (kernels), which are likely to belong to the same true cluster. CLICK (Cluster Identification via Connectivity Kernels) is an example of this technique [Sharon and Shamir, 2000]. The CLICK algorithm first preprocesses the input data and forms a weighted similarity graph. Recursively it then partitions the current set of elements into two subsets. Before a partition, the algorithm tests if the subgraph induced by the current subset of elements is a kernel of a cluster. If it is a kernel, the subgraph is not partitioned further. Otherwise, a minimum weight cut is computed in the subgraph, and the current set of elements is split into the two subsets separated by that cut. The output is a list of kernels which serve as a basis for the eventual clusters [Sharon and Shamir, 2000].Graphtheoretic approaches were also used for the related problem of clustering cDNAs based on their oligonucleotide fingerprints [Hartuv et al., 1999 and Lennon and Lehrach, 1991].

Recognizing that genes are not randomly distributed, Turkheimer et al. (2006) created a mathematical model (CHROMOWAVE) that is based on single value decomposition and Haar wavelets. It detects variations in expression of spatially related genes and visualizes them on chromosomes.

Many approaches try to use reverse engineering to construct genetic networks from gene expression data. Boolean networks [Kauffman, 1969] are a simple way to represent the interaction between genes. Each gene is considered a binary variable (ON or OFF) that are regulated by other genes through Boolean functions [D'haeseleer et al., 2000]. These networks provide a conceptual tool for investigating network organization, but there are many criticisms on using the Boolean network approach [Akutsu et al., 2000].

Other approaches have used linear differential equations to reconstruct genetic networks [Chen et al., 1999]. In Chen et al. (1999), they used a linear transcription model for gene expression, as well as two algorithms (minimum weight solutions to linear equations and Fourier transform for stable systems) to construct the model from a set of temporal samples of mRNAs and proteins. With this model, they were able to determine most of the gene regulation in the genome level from a minor set of accurate temporal data.

Bayesian networks are also used to reverse engineer networks. These networks represent the dependence structure between expression levels of different genes. Friedman et al. (2000) used a technique called the Sparse Candidate algorithm [Friedman, Nachman and Pe'er 1999] to identify a small number of candidate parents for each gene. This restricts the search for networks in which only the candidate parents of a variable can be its parents, thus making the search space for finding a good network structure. In order to specify a Bayesian model, the type of local probability models that are learned need to be selected. Friedman et al. (2000) used two different models (multinomial and linear Gaussian), and they were able to predict sensible relations between genes of known function. Chen et al. (2006) developed a methodology that processes the expression data to find a suitable node ordering that can be used by the greedy search algorithm, K2 algorithm [Copper and Herskovits, 1992], to learn the structure of Bayesian
networks. This method can identify networks that are close to the optimal structures [Chen et al., 2006].

Little effort has been placed on analyzing the genome for these relationships before initiating the microarray experiments. My research focuses on developing software to help design (i.e. which probe sets should be included on the microarray chip) the actual chips. The software designed here can provide researchers with a way to focus their search for coregulated genes, which will lead to better chip design, and ultimately save the researcher time and money. This work could be pivotal for research by providing a more focused experimental approach to elucidate biological roles and relationships between genes.

### 2.6.2 Determining Functionality

Gene function can be established using sequence information. Comparing the sequence of the gene with the unknown functionality with known genes in a database (such as GenBank at the National Center for Biotechnology Information or NCBI) is a technique for establishing possible functionality of the unknown. Sequences are compared using algorithms like BLAST (Basic Local Alignment Search Tool) [Altschul et al., 1990]. The premise of these searches is that if two sequences are similar, they probably share a common evolutionary ancestor (i.e. they are homologues) and possibly the same functionality. For example, if your search results contained a lot of sequences that showed a high degree of similarity with your given gene and that these similar genes belonged to the same family of genes that were known to perform "function x " then you could conclude that your given gene encoded for a gene product with a similar function. Protein motifs can also be searched for to help determine functionality. A motif is a distinctive pattern of amino acids that is conserved across many proteins, and it gives a particular function to the protein ["Determining Gene..."].

When there are no homologues for the given gene, a biological approach exists to assist in determining its function. This approach involves creating a mutation in that gene and then observing the effects of the mutation on the organism to determine the genes function. This is called a knockout study, and these studies give researchers insight to how genes work in cells. The problem with some knockout studies is that the defects resulting from such experiments can yield no apparent changes because other genes can fulfill that same function as the altered gene. Other genetic manipulations can be lethal early in development, eliminating the possibility of observing any changes at all.

The work done in my research helps start limiting the list of possible functions of a gene product by using information that is known about the gene's neighbors. It does not replace using knockouts to establish functionality. The suggestions made by the software are to be used as guides.

### 2.6.3 Establishing Similarity Using Text

Establishing similarity between proteins/gene products is usually done using sequence alignment and/or structural alignment. Similarity is a quantitative term that defines the degree of sequence match between two compared sequences. Other methods have expanded the definition of similarity that go beyond mere alignment. These methods investigate establishing relationships between proteins based on the functionality of the proteins. Marcotte et al. showed that general biochemical functions of proteins can be inferred by associating proteins on the basis of properties other than the similarity between their amino acid sequences. These properties associate proteins that are functionally related. Proteins that are functionally related participate in a common structural complex, metabolic pathway, biological process or closely related physiological function. Functional relationships can provide information that can go beyond the capabilities of traditional sequence matching [Marcotte et al., 1999].

Relationships between proteins can also be established based on text. Wren et al. identified one of the best sources to identify relationships, commonalities, and functional groupings among a heterogeneous set of objects (e.g. genes, phenotypes, chemicals, diseases) is the use of scientific literature. The technique they described enables object sets to be scored for their cohesiveness, as judged by their co-occurrence within scientific literature (MEDLINE). This method allows other objects to be identified and evaluated for their potential cohesiveness or relevance in relation to the analyzed set (cohesiveness is the metric used - it enables researchers to infer that an experimental grouping is purposeful assuming the grouped objects are adequately represented within the literature). Prior use of term co-occurrence has been used to find tentative relationships between objects such as genes [Stapley and Benoit, 2000; Jenssen et al., 2001], proteins [Blaschke et al., 1999] and drugs [Rindflesch et al., 2000]. Wren et al. noted that cooccurrence does not necessarily reflect the existence of a meaningful relationship, so Wren et al. used Fuzzy Set Theory to assign a weight to the relatedness of two objects based upon their frequency of co-occurrence. Ambiguous gene names are noted to be problematic in cooccurrence networks [Wren et al., 2004].

Determining similar function among proteins has been explored. Altman and Raychaudhuri (2001) showed that there is an interest in using information from genes with a common function to search for additional genes that share this function. They stated that if the function of a gene is unknown, one can analyze microarray data and make an educated guess as to the function of the gene through "guilt by association". Guilt by association means that if a non-annotated gene is regulated in a similar fashion compared to a group of annotated genes, then it has a similar function to the annotated genes. The co-expression of genes may imply that they share common regulatory mechanisms. There is controversy about this notion because regulatory mechanisms can be mixed and combined in ways that could lead to both convergent regulation (similar
temporal expression patterns, different control strategies) and divergent regulation (similar control regions, put together in such a way that the effect on expression is different) [Altman and Raychaudhuri, 2001].

These studies suggest that there are alternative ways to establish similarity other than sequence or structural alignment. My research utilizes a search engine to retrieve similar genomic features to suggest possible locations to look for other features that may be involved in the same network. Again, this is trying to provide a more focused experimental approach to elucidate biological roles and relationships between genes.

## 3. Methodology

### 3.1 Approach

From previous research (See Background), the location of a genomic feature and its surrounding neighborhood have an influence on the feature's network involvement and its functionality. Determining what attributes of the surrounding neighborhood had the biggest impact on classifying a feature to the correct network and functionality needed to be established. Different neighborhood attributes were tested and different sizes of neighborhoods were examined with different classifiers. Once the best neighborhood attributes, the best neighborhood size and better classifier were determined, software could be developed to suggest locations of interest for network involvement and possible functionality for a given gene.

A feature's neighborhood attributes were determined by the attributes of its nearest neighbors on both strands and both upstream and downstream to the feature. The neighbor attributes that remained constant for the experiments were: the neighbor's number (indicating how close it was to the feature, one being the closest), the neighbor's strand and the neighbor's distance (determined from the mid-point of each feature). The different attributes that were then tested separately in classifying network involvement for a feature were: the neighbor's SGDID, the neighbor's Gene Ontology information and the neighbor's network involvement. The different attributes that were then tested separately in classifying functionality for a feature were: the neighbor's SGDID and the neighbor's Gene Ontology information.

The size of the neighborhood was determined by how many nearest neighbors should be included. All experiments were repeated using different sizes of neighborhoods. These neighborhoods included information from the nearest neighbors, the two-nearest neighbors,
the five-nearest neighbors and the ten-nearest neighbors of each feature on both strands and both upstream and downstream from the feature.

The Saccharomyces cerevisiae genome was selected for classification because of all the complete genomes, it has the most ideal characteristics for a test case. The Saccharomyces cerevisiae genome shows a high amount clustering of genes that are involved in the same metabolic pathway [Lee and Sonnhammer, 2003] and it shows clustering of essential genes into regions of low recombination [Pál and Hurst, 2003]. Incidences of highly coordinated expression of linked genes have also been found in yeast [Cohen et al., 2000 and Kruglyak, 2000]. Clusters of coexpressed genes in yeast rarely exceed ten genes or a few kilobases, so larger clusters were not examined [Hurst et al., 2004].

In this study, various classification techniques (KStar, MultiBoostAB using the J48 classifier and Naïve Bayes) were used on genomic data of Saccharomyces cerevisiae. All of the classifiers that were used are contained in the software package, WEKA [Frank and Witten, 2005], and this software package was used setting the Java maximum heap size set to 12 GB.

Experiments were performed to determine what classifier yielded the best results and:

1. What neighbor information is the most beneficial in classifying genomic features to the correct network, and how many neighbors should be included.
a. Partitioning the classification process by chromosome was also investigated.
2. What neighbor information is the most beneficial in classifying genomic features to the correct Gene Ontology term, and how many neighbors should be included.
a. Partitioning the classification process by chromosome was also investigated.

Additional experiments were performed to determine if locations of interest could be suggested based on a description of a feature. Experiments were performed to determine:
3. Whether genomic features can be classified to the correct location
a. Classification experiments were performed to see if a feature's chromosome could be suggested based on either its network involvement or its Gene Ontology term.
b. Classification experiments were performed to see if a feature's start position could be suggested based on its chromosome and either its network involvement or its Gene Ontology term.
4. Whether given a textual description of a feature, would finding the networks that the top ten similar features (similarity based on IR) were involved in be helpful by suggesting looking by the feature's involved in these networks as areas of possible interest.

Based on the results from the experiments mentioned in numbers 1 and 2 above, it was determined that the MultiBoosting algorithm using the J48 classifier and that partitioning the classification process by chromosome yielded the better classification of features to their networks and to their Gene Ontology terms. It was also determined that a neighborhood that included each feature's two-nearest neighbors and the neighbors' Gene Ontology information provided the best information for these classifications.

The decision trees generated from these experiments were then converted into if-then rules to be utilized by two different rule-based systems. The first rule-based system, when given the location of a genomic feature of interest, will suggest areas in the genome to look for other genomic features that may interact with the given feature (See Figure 3.1). The other system, when given the location of a genomic feature of interest, suggests how its gene product(s)
behave in a cellular context (Gene Ontology term) (See Figure 3.2). A detailed discussion of how the systems work will be discussed in section 3.2.

Based on the results from the experiments mentioned in number 3, it was determined that if either a feature's network involvement or its Gene Ontology term is known, it would be hard to determine accurately which chromosome it is located on. Even if the chromosome number is determined, classifying the feature to the correct start position given either its network involvement or its Gene Ontology term is still hard to do. A Nearest-neighbor-like algorithm using non-nested generalized exemplars did obtain rules that only suggested some of the locations, not all. These rules were adapted (See Appendix G) to create a rule-based system that suggested areas of interest based on a given network.

Based on the results from the experiments mentioned in number 4, it was determined that textual descriptions are helpful. A search engine based on genomic feature descriptions was created. The search engine was developed to take as input a description of a feature. Based on the network involvement of the top ten most similar features to the given feature, locations of start positions in the genome for other features that may participate in a network with the given feature are provided. These locations are suggested using rules adapted from experiments mentioned in number 3. A diagram of the process can be seen in Figure 3.2.3, and a description of the system will be discussed in section 3.2.

### 3.2 Systems

Different systems were designed to utilize the knowledge gained in the experiments discussed in the previous section. The first rule-based system suggests areas in the genome to look for other genomic features that may interact with a given feature when given its location (See Figure 3.1). The user must provide the system with the chromosome number, the strand and the start
position of a feature (box 1). The system then finds the two-nearest neighbors (both upstream and downstream, and on both strands) of that feature (box 2). It then looks up the Gene Ontology term (if known) for each neighbor and determines their neighbor numbers (box 3). Information from each neighbor is then ran through the rule-based system adapted from the experiments mentioned in section 3.1 number 1 (box 4). Network suggestions are then made by the rule-based system (box 5) and given to two different methods. The first method looks up all the features that are known to participate in that network (box 6a) and reports their locations, so the user can focus their search around these areas (box 7a). The second method relies on the rules adapted from the experiments mentioned in section 3.1 number 3 . The rule-based system (box 6b) takes the network suggestions (box 5) and suggests areas of interest where features that may be involved with the given feature are located (box 7b).


Figure 3.1. Diagram of software designed to suggest areas of interest for networks.

The other rule-based system suggests a gene product's Gene Ontology term when given its location (See Figure 3.2). The user must provide the system with the chromosome number, the strand and the start position of a feature (box 1). The system then finds the two-nearest neighbors (both upstream and downstream, and on both strands) of that feature (box 2). It then looks up the Gene Ontology term (if known) for each neighbor and determines their neighbor numbers (box 3). Information from each neighbor is then ran through the rule-based system adapted from the experiments mentioned in section 3.1 number 2 (box 4). The system then generates a list of Gene Ontology terms (box 5) to suggest to the user.


Figure 3.2. Diagram of software designed to suggest GO Terms.

The last system uses a search engine to recommend areas of interest for network involvement. A search engine was generated using the textual descriptions provided for features at the SGD website (See Figure 3.3). Each feature description was processed and the standard dictionary and
posting files were created. The user provides a textual description of the feature, that description is processed in the same manner the features in the search engine were processed (box 1). The description is fed into the search engine (box 2) and the top ten most similar features are returned (box 3). For each of these features, their network involvement (if known) is looked up (box 4). The network involvement of the top ten are then fed into the rule-based system that was created based on experiments mentioned in section 3.1 number 3. The rule-based system (box 6) takes the network suggestions (box 5) and suggests areas of interest where features that may be involved with the given feature are located (box 7).


Figure 3.3. Diagram of how location suggestion is made.

These tools differ from current approaches, because they are using what is already known to help guide future work. For network involvement, little effort has been placed on developing software to help design (i.e. which probe sets should be included on the microarray chip) the
actual chips. The software designed here can provide researchers with a way to focus their search for coregulated genes, which will lead to better chip design, and ultimately save the researcher time and money. This work could be pivotal for research by providing a more focused experimental approach to elucidate biological roles and relationships between genes. The user need only to provide a location or a textual description of a feature for my software to generate areas of interest in the genome.

The work done in this study helps start limiting the list of possible functions of a gene product by using information that is known about the gene's neighbors. It does not replace using knockouts to establish functionality. The suggestions made by the software are to be used as guides.

## 4 Experiments

### 4.1 Description of Experiments

### 4.1.1. Classifying Features to Networks

Experiments were performed to establish if classification algorithms can correctly classify network involvement (based on KEGG) of a genomic feature (a feature being: gene, ARS, etc.). The classification algorithms used include KStar, MultiBoosting the J48 classifier and Naïve Bayes. Different neighborhood attributes were used in the classification process to determine what information is the most beneficial in generating rules for classifying features to their correct network. Different neighborhood sizes were examined to determine how many nearest neighbors should be included around each feature to provide better classification rules. As a baseline, the classification process was performed using no neighbor information. Also, the classification process was performed using the entire genome, and then it was repeated on each chromosome to examine if partitioning the classification process would yield better rules. The classification process was not repeated on the mitochondrial chromosome because it only has one known network on it.

The classifiers were trained using different neighborhood attributes. A feature's neighborhood attributes were determined by the attributes of its nearest neighbors on both strands and both upstream and downstream to the feature. The feature's attributes that remained constant in the training process were: chromosome number, start position, stop position and strand. The neighbor attributes that remained constant for the training process were: the neighbor's number (indicating how close it was to the feature, one being the closest), the neighbor's strand and the neighbor's distance (determined from the mid-point of each feature). The different attributes that were then tested separately in classifying network involvement for a feature were: the
neighbor's SGDID, the neighbor's Gene Ontology information and the neighbor's network involvement.

The size of the neighborhood was determined by how many nearest neighbors should be included. All experiments were repeated using different sizes of neighborhoods. These neighborhoods included information from the nearest neighbors, the two-nearest neighbors, the five-nearest neighbors and the ten-nearest neighbors of each feature on both strands and both upstream and downstream from the feature. Certain sized neighborhoods were excluded only when the heap size of the Java Virtual Machine was exceeded.

### 4.1.2. Classifying features to Gene Ontology terms

Experiments were performed to establish if classification algorithms can correctly classify Gene Ontology terms of a genomic feature. The classification algorithms used include KStar, MultiBoosting using the J48 classifier and Naïve Bayes. Different neighborhood attributes were used in the classification process to determine what information is the most beneficial in generating rules for classifying features to their correct network. Different neighborhood sizes were examined to determine how many nearest neighbors should be included around each feature to provide better classification rules. As a baseline, the classification process was performed using no neighbor information. Also, the classification process was performed using the entire genome, and then it was repeated on each chromosome to examine if partitioning the classification process would yield better rules.

The classifiers were trained using different neighborhood attributes. A feature's neighborhood attributes were determined by the attributes of its nearest neighbors on both strands and both upstream and downstream to the feature. The feature's attributes that remained constant in the
training process were: chromosome number, start position, stop position and strand. The neighbor attributes that remained constant for the training process were: the neighbor's number (indicating how close it was to the feature, one being the closest), the neighbor's strand and the neighbor's distance (determined from the mid-point of each feature). The different attributes that were then tested separately in classifying network involvement for a feature were: the neighbor's SGDID and the neighbor's Gene Ontology information.

The size of the neighborhood was determined by how many nearest neighbors should be included. All experiments were repeated using different sizes of neighborhoods. These neighborhoods included information from the nearest neighbors, the two-nearest neighbors, the five-nearest neighbors and the ten-nearest neighbors of each feature on both strands and both upstream and downstream from the feature. The exclusion of including each feature's tennearest neighbors only occurred when the heap size of the Java Virtual Machine was exceeded.

### 4.1.3 Suggesting Locations of Interest

Experiments were performed to see if classification algorithms could correctly classify the chromosome of a genomic feature and the start position of a genomic feature. The classification algorithms used include KStar, MultiBoosting using the J48 classifier, Naïve Bayes, and a Nearest-neighbor-like algorithm using non-nested generalized exemplars. Different feature attributes were used in the classification process to determine what information would generate better rules for classifying features to their correct chromosome and their correct start position. The goal was to produce rules that would suggest areas in the genome that contained features with a certain quality (belonged to a certain network or Gene Ontology term).

When training classifiers to classify features to the correct chromosome, different feature attributes were used to determine what attributes produced better rules. Classifiers were given training data that included the feature's chromosome number and its network involvement. Classifiers were also given training data that included the feature's chromosome number and its Gene Ontology term.

When training classifiers to classify a features to the start position, different feature attributes were used to determine what attributes produced the better rules. Classifiers were given training data that included the feature's chromosome number, its discretized start position, its strand and its network involvement. Classifiers were also given training data that included the feature's chromosome number, its discretized start position, its strand and its Gene Ontology term.

Experiments were also performed to determine if textual descriptions of features could be used to suggest areas of interest in the genome for network involvement. It was examined if retrieving the network involvement of the top ten most similar features (similarity based on IR) of a given feature would be helpful in suggesting where to look for other features that may be involved with that given feature.

### 4.2 Datasets

For these experiments, the genome feature information was taken from the Saccharomyces Genome Database (http://www.yeastgenome.org/). The files were obtained from the Anonymous FTP site. Information for all genomic features that have been physically mapped was used. These features include genes, telomeric elements, autonomously replicating sequences, hypothetical proteins, putative proteins of unknown function, elements of dubious open reading frames and centromeres. The files were formatted as CSV (comma-separated values) files for processing by WEKA.

For each feature, its primary information was used. This information includes the feature's chromosome number (one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, fifteen, sixteen, mito), start position (in bp), stop position (in bp) and strand ("W" for Watson and "C" for Crick). Depending on what experiment was being run, additional information was included. If information was not available, it was substituted with a "?".

For experiments that pertained to classifying networks, for each network that a feature is involved in, an entry into the file that was analyzed was added. For example, for a feature that belongs to two different networks, sce00910 (nitrogen metabolism) and sce00251 (glutamate metabolism), it will have two different entries into the file. See Table 4.1 for an example.

| Chromosome | Start | Stop | Strand | Network |
| :--- | :--- | :--- | :--- | :--- |
| one | 31568 | 32941 | W | sce00910 |
| one | 31568 | 32941 | W | sce00251 |

Table 4.1. Shows the format of data files that included network involvement but did not include neighbor information

For experiments that analyzed how neighbors influenced the classification of networks, additional information was added. The information that was constant in all neighbor experiments included the neighbor number ( 1 being closest), the neighbor's strand, and the distance from the feature's midpoint to the neighbor's midpoint.

For experiments that investigated the influence of neighbor's SGDID (Saccharomyces Genome Database identification) on the feature's network classification, the neighbor's SGDID was included. For each of the feature's neighbors, a separate entry was created for each of the feature's different networks. See Table 4.2 for an example.

| Chromosome | Start | Stop | Strand | Neigh- <br> bor <br> Number | Neigh- <br> bor <br> Strand | Distance | Neighbor <br> SDGID | Network |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| one | 31568 | 32941 | W | 1 | W | 1189 | S000118317 | sce00910 |
| one | 31568 | 32941 | W | 1 | W | 1189 | S000118317 | sce00251 |
| one | 31568 | 32941 | W | 2 | W | 9861 | S000006787 | sce00910 |
| one | 31568 | 32941 | W | 2 | W | 9861 | S000006787 | sce00251 |

Table 4.2. Shows the format of the data files that included neighbor's SGDID and feature's network involvement.

For experiments that investigated the influence of neighbor's Gene Ontology term on the feature's network classification, the neighbor's Gene Ontology term was included. For each of the neighbor's Gene Ontology terms, a separate entry was created for each of the feature's different networks. See Table 4.3 for an example.

| Chromosome | Start | Stop | Strand | Neighbor <br> Number | Neighbor <br> Strand | Distance | Neighbor <br> GO Term | Network |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| one | 45023 | 42882 | C | 1 | W | 3123 | cytoplasm | sce00010 |
| one | 45023 | 42882 | C | 1 | W | 3123 | cytoplasm | sce00620 |
| one | 45023 | 42882 | C | 1 | W | 3123 | cytoplasm | sce00720 |
| one | 45023 | 42882 | C | 1 | W | 3123 | endoplasmic <br> reticulum | sce00010 |
| one | 45023 | 42882 | C | 1 | W | 3123 | endoplasmic <br> reticulum | sce00620 |
| one | 45023 | 42882 | C | 1 | W | 3123 | endoplasmic <br> reticulum | sce00720 |
| one | 45023 | 42882 | C | 1 | W | 3123 | transporter <br> activity | sce00010 |
| one | 45023 | 42882 | C | 1 | W | 3123 | transporter <br> activity | sce00620 |
| one | 45023 | 42882 | C | 1 | W | 3123 | transporter <br> activity | sce00720 |

Table 4.3. Shows the format of the data files that included neighbor's Gene Ontology term and feature's network involvement.

For experiments that investigated the influence of neighbor's network involvement on the feature's network classification, the neighbor's network involvement was included. For each of the networks a neighbor was involved in, a separate entry was created for each of the feature's different networks. See Table 4.4 for an example.

| Chromosome | Start | Stop | Strand | Neighbor <br> Number | Neighbor <br> Strand | Distance | Neighbor <br> Network | Network |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| one | 69526 | 68717 | C | 1 | W | 4998 | $?$ | sce00860 |
| one | 69526 | 68717 | C | 2 | W | 4998 | sce04010 | sce00860 |
| one | 69526 | 68717 | C | 1 | C | 2471 | $?$ | sce00860 |
| one | 69526 | 68717 | C | 2 | C | 2471 | sce04111 | sce00860 |
| one | 69526 | 68717 | C | 1 | W | 1253 | $?$ | sce00860 |
| one | 69526 | 68717 | C | 2 | W | 3417 | sce00010 | sce00860 |
| one | 69526 | 68717 | C | 2 | W | 3417 | sce00230 | sce00860 |
| one | 69526 | 68717 | C | 2 | W | 3417 | sce00620 | sce00860 |
| one | 69526 | 68717 | C | 2 | W | 3417 | sce00710 | sce00860 |

Table 4.4. Shows the format of the data files that included neighbor's network and feature's network involvement.

For experiments that pertained to classifying Gene Ontology terms, for each Gene Ontology term that a feature has, an entry into the file that was analyzed was added. For example, if a feature has three different Gene Ontology terms, cellular component, molecular function and ribosome biogenesis and assembly, it will have three different entries into the file. See Table 4.5 for an example.

| Chromosome | Start | Stop | Strand | GO aspect | GO term |
| :--- | :--- | :--- | :--- | :--- | :--- |
| one | 99306 | 99869 | W | C | cellular component |
| one | 99306 | 99869 | W | F | molecular function |
| one | 99306 | 99869 | W | P | ribosome biogenesis and assembly |

Table 4.5. Shows the format of data files that included Gene Ontology information but did not include neighbor information

For experiments that investigated the influence of neighbor's SGDID (Saccharomyces Genome Database identification) on the feature's Gene Ontology term classification, the neighbor's SGDID was included. For each of the feature's neighbors, a separate entry was created for each of the feature's different Gene Ontology terms. See Table 4.6 for an example.

| Chromosome | Start | Stop | Strand | Neigh- <br> bor <br> Number | Neigh <br> bor <br> Strand | Neighbor <br> SGDID | Distance | GO term |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| one | 99306 | 99869 | W | 1 | W | S000033629 | 4507 | cellular <br> com- <br> ponent |
| one | 99306 | 99869 | W | 1 | W | S000033629 | 4507 | molecular <br> function |
| one | 99306 | 99869 | W | 1 | W | S000033629 | 4507 | ribosome <br> biogenesis <br> and <br> assembly |
| one | 99306 | 99869 | W | 2 | W | S000000025 | 4507 | cellular <br> com- <br> ponent |
| one | 99306 | 99869 | W | 2 | W | S000000025 | 4507 | molecular <br> function |
| one | 99306 | 99869 | W | 2 | W | S000000025 | 4507 | ribosome <br> biogenesis <br> and <br> assembly |

Table 4.6. Shows the format of the data files that included neighbor's SGDID and feature's Gene Ontology information.

For experiments that investigated the influence of neighbor's Gene Ontology information on the feature's Gene Ontology term classification, the neighbor's Gene Ontology information was included. For each of the neighbor's Gene Ontology terms, a separate entry was created for each of the feature's different Gene Ontology terms. See Table 4.7 for an example.

| Chromo -some | Start | Stop | Strand | Neigh- <br> bor <br> num- <br> ber | Neigh -bor strand | Distance | Neig h-bor GO aspect | Neigh- <br> bor GO <br> term | GO aspect | GO term |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| one | 2169 | 1807 | C | 1 | W | 605 | C | cellular componen t | C | cellular component |
| one | 2169 | 1807 | C | 1 | W | 605 | C | cellular <br> componen <br> t | F | molecular function |
| one | 2169 | 1807 | C | 1 | W | 605 | C | cellular componen t | P | sporulation |
| one | 2169 | 1807 | C | 1 | W | 605 | F | molecular function | C | cellular component |
| one | 2169 | 1807 | C | 1 | W | 605 | F | molecular function | F | molecular function |
| one | 2169 | 1807 | C | 1 | W | 605 | F | molecular function | P | sporulation |
| one | 2169 | 1807 | C | 1 | W | 605 | P | biological process | C | cellular component |
| one | 2169 | 1807 | C | 1 | W | 605 | P | biological process | F | molecular function |
| one | 2169 | 1807 | C | 1 | W | 605 | P | biological process | P | sporulation |
| one | 2169 | 1807 | C | 2 | W | 605 | ? | ? | C | cellular component |
| one | 2169 | 1807 | C | 2 | W | 605 | ? | ? | F | molecular function |
| one | 2169 | 1807 | C | 2 | W | 605 | ? | ? | P | sporulation |

Table 4.7. Shows the format of the data files that included neighbor's Gene Ontology information and feature's Gene Ontology information.

### 4.3 Algorithms

All datasets were analyzed by WEKA. The algorithms used to analyze the datasets are Naïve Bayes, KStar and MultiBoostAB using the J48 classifier. An exhaustive analysis of all possible algorithms was not performed. The performance of a classifier is dataset dependent. We still do not have a single classifier that can reliably outperform all others on a given dataset [van der Walt and Barnard, 2006]. The process of selecting a classifier is a process of trial and error, so a sampling of algorithms from different families of classification was examined.

The Naïve Bayesian classifier was selected because often the simplest approach is the best, and studies that have compared various classification algorithms found the Naïve Bayesian classifier to be comparable in performance with classification trees and with neural network classifiers [Stewart, 2002]. The KStar classifier was selected because it tends to perform better than other lazy classifiers (IB1-5 [Aha et al., 1991]), and in some cases it performs better than Quinlan's C4.5 [Cleary and Trigg, 1995]. J48 (Java version of C4.5) was selected because even though there are other tree classifiers, Quinlan's research has always been on the "forefront of decision tree induction" and it is an algorithm that is well studied [Witten and Frank, 2005]. MultiBoosting was used in conjunction with J48 to produce decision committees with lower error than either AdaBoost or wagging over a broad array of datasets [Webb, 2000].

For the KStar experiments, a default blend setting of $20 \%$ was used because that is the value "which seems to work well for most datasets" [Cleary and Trigg, 1995]. The missing values were replaced with the average of the entropy curve for that attribute in which the missing value belonged.

For the MultiBoostAB experiments, reweighting was used instead of resampling because past experiments produced results the suggested "reweighting is more effective than resampling" [Webb, 2000]. The other settings were set to default values.

For the experiments that required IR techniques, 8287 genomic feature descriptions were obtained from the SGD. These descriptions were pre-processed for IR retrieval. The stop-list that was used can be found in Appendix A.

### 4.4 Metrics

The performance of the classification process can be evaluated by several metrics. In this study, six metrics were focused on: percentage of correctly classified instances, the Kappa statistic, precision, recall, the F-measure and the area under the ROC (receiver operating characteristics) curve.

The percentage of correctly classified instances is a basic accuracy measurement that can be determined by the following:

$$
\text { Percent Correctly Classified }=\frac{\text { number of correctly classified instances }}{\text { number of instances }}
$$

The Kappa statistic measures the agreement between predicted and observed categorization. As a test statistic, Kappa can verify that agreement exceeds chance levels. It has possible values that range from +1 (perfect agreement) to 0 (no agreement above that expected by chance, or in other words the agreement is equal to chance) and then to -1 (perfect disagreement). When

Kappa is negative, it means that the agreement is less than what would be expected by chance. It is rare that to get perfect agreement. Kappa is calculated as:

$$
\varkappa=\frac{P(A)-P(E)}{1-P(E)}
$$

where $P(A)$ is the observed agreement (between the classifier and the actual truth), and $P(E)$ is the chance agreement.

There are different interpretations as to what is a good level of agreement. One possible interpretation of Kappa provided by Altman (1991) can found in Table 4.8.

| Agreement | Range of values |
| :--- | :--- |
| Poor | Less than 0.20 |
| Fair | 0.20 to 0.40 |
| Moderate | 0.40 to 0.60 |
| Good | 0.60 to 0.80 |
| Very Good | 0.80 to 1.00 |

Table 4.8 Interpretation of Kappa statistic.

The precision and recall scores are similar to those used in IR. In the classification process, a precision score of 1.0 for a class $X$ means that every item labeled as belonging to class $X$ does indeed belong to class $X$ (but says nothing about the number of items from class $X$ that were not labeled correctly). In the classification process, a recall score of 1.0 means that every item from class $X$ was labeled as belonging to class $X$ (but says nothing about how many other items were incorrectly also labeled as belonging to class $X$ ). Precision and recall are calculated in the following manner:

# $\operatorname{Precision}(X)=\frac{\text { number of correctly classified instances of class } \mathrm{X}}{\text { number of instances classified as belonging to class } \mathrm{X}}$ 

$$
\operatorname{Recall}(X)=\frac{\text { number of correctly classified instances of class } X}{\text { number of instances in class }}
$$

Often, there is an inverse relationship between precision and recall, where it is possible to increase one at the cost of reducing the other. For example, in a classification system for deciding whether or not, say, a fruit is a tomato, can achieve high precision by only classifying fruits with the exact right shape (round) and color (red) as tomatoes, but at the cost of low recall due to the number of false negatives from tomatoes that did not quite match the specification, like a green Roma tomato.

Precision and recall are rarely discussed in isolation. They can be combined into a single measure, such as the F-measure. The F-measure is a weighted harmonic mean of precision and recall. It is calculated in the following manner:

$$
F=2 * \frac{(\text { precision } * \text { recall })}{(\text { precision }+ \text { recall })}
$$

In statistics, the F-measure is a measure of a test's accuracy. The F-measure can be interpreted as a weighted average of the precision and recall, where an F-measure reaches its best value at 1.0 and worst score at 0 .

Another way to measure accuracy is to use the area under the ROC curve. The area has a nice interpretation as the probability that the classifier ranks a randomly chosen positive instance above a randomly chosen negative one. The accuracy of the test depends on how well the classifier separates the group being tested into those that belong to a class and those that do not belong to a class.

An area of 1.0 represents a perfect test. An area of 0.5 represents a worthless test. For better understanding of the ROC curves, Thomas G. Tape provides a graph of some curves with an interpretation as follows:


Figure 4.1. Taken from Interpreting Diagnostic Tests web page by Thomas G. Tape to illustrate the interpretation of the area under the ROC curve.

A rough guide for classifying the accuracy is provided by Tape (See Table 4.9).

| Results | Area ranges |
| :--- | :--- |
| Fail | $0.50-0.60$ |
| Poor | $0.60-0.70$ |
| Fair | $0.70-0.80$ |
| Good | $0.80-0.90$ |
| Excellent | $0.90-1.0$ |

Table 4.9. Interpretation of area under ROC curve from Interpreting Diagnostic Tests web page by Thomas G. Tape.

## 5 Results

### 5.1 Classifying Genomic Features to Networks

The results of experiments that determined if altering the type of neighbor information given the classifiers altered the accuracy of the classification process follow.

### 5.1.1 Using Neighbor's SGDID

A series of experiments was performed to determine if given the properties of chromosome number, start position, stop position, strand, neighbor number, neighbor SGDID, neighbor strand, neighbor distance (determined from the mid-point of each feature), and network involvement information from KEGG for genomic features (features being: genes, ARS, etc) could classification algorithms correctly classify network involvement. This was also performed including information from the nearest, two-, five- and ten- nearest neighbors of each feature. This was also repeated on the individual chromosomes. The mitochondrial chromosome is excluded. Having only one known network on the mitochondrial chromosome, classifiers cannot classify unary classes.

The results of the experiments are based on averages of ten different runs of each dataset being randomized and then split for training and testing. The training sets were set to contain $66 \%$ of the instances randomly selected from each dataset.

### 5.1.1.1 Results for Entire Genome

When using only a genomic feature's chromosome number, start position, stop position, strand, and network involvement without neighbor information, we see that the percentage of genomic features correctly classified to the correct network does not vary greatly when different classifiers are used (See Figure 5.1). When no neighbor information is included, less than $20 \%$ of the
genomic features are classified to the correct network regardless of the classifier used. As more neighbor information is included, the percentage of correctly classified features increases for the KStar classifier. For the MultiBoostAB algorithm using the J48 classifier, the percentage of correctly classified features decreases with the inclusion of neighbor information. When the Naïve Bayesian classifier is used, the percentage of correctly classified genomic features decreases with the inclusion of neighbor information until information from each feature's ten-nearest neighbors is included where it then increases.


Figure 5.1. Percent of genomic features classified for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

The agreement between the predicted and observed classification of the genomic features can be seen in the Kappa statistic. When no neighbor information is included, all of the classifiers have poor agreement between the predicted and observed classification (See Figure 5.2). The KStar classifier achieves moderate agreement as more neighbor information is added. A moderate agreement is achieved with the Naïve Bayesian classifier when information on the ten-nearest neighbors of a genomic feature is included. The MultiBoostAB algorithm using the J48 classifier
never achieves an agreement above poor. In fact, the inclusion of neighbor information decreases the agreement between predicted and observed classification.


Figure 5.2. Kappa statistic for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

The overall precision of the classification process is low for all classifiers (See Figure 5.3). The KStar classifier obtains a precision score above 0.2 when information from each feature's twonearest neighbors is included. In all other instances, the precision scores are never above 0.15.


Figure 5.3. Precision for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

The recall scores of the classifiers mimic the precision scores the classifiers obtained (See Figure 5.4). The highest recall score obtained by the KStar classifier is when information from each feature's two-nearest neighbors is included. The Naïve Bayesian classifier does not obtain a recall score above 0.12 even after information from each feature's ten-nearest neighbors is included. For the MultiBoostAB algorithm using the J48 classifier, the recall scores decrease with the inclusion of neighbor information.


Figure 5.4. Recall for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

Not surprisingly, the F-measures obtained by the different classifiers mimic their precision and recall scores (See Figure 5.5). For the KStar classifier, the highest F-measure is obtained when information from each feature's two-nearest neighbors is included. There is not a dramatic change in F-measures for the Naïve Bayesian classifier until information about the ten-nearest neighbors for each feature is included. For the MultiBoostAB algorithm using the J 48 classifier, the F-measure decreases with the inclusion of neighbor information.


Figure 5.5. F-measure for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

When no neighbor information is included, the MultiBoostAB algorithm using the J48 classifier is the only classifier that obtains an accuracy score, according to the area under the ROC curve, that is not failing, but it is considered poor (See Figure 5.6). The accuracy does not improve when neighbor information is included. The KStar classifier improves its accuracy to excellent when information about the neighbor information is included, but the accuracy drops to failing when information from each feature's ten-nearest neighbors is included. With the Naïve Bayesian classifier, the accuracy of the classification improves to fair when information from each feature's five-nearest neighbors is included and then improves to excellent when information from the ten-nearest neighbors is included.


Figure 5.6. Area under ROC curve for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

To see the effect on the accuracy of the KStar classification for each network, the F-measures for each network were analyzed. The table of the F-measures can be found in Appendix B. The addition of neighbor information improves the accuracy of the classification process for most networks (See Figures 5.7-5.9).

Some networks attain F-measures that are above 0.6 with no neighbor information included implying that the location of these features alone is indicative of their network involvement. These networks are the ribosome network and the polyunsaturated fatty acid biosynthesis network. Other networks (the oxidative phosphorylation network, the SNARE interactions in vesicular transport network, the ubiquitin biosynthesis network, the proteasome network, the valine, leucine and isoleucine biosynthesis network, the glycerophospholipid metabolism network, the cell cycle network, the aminoacyl-tRNA biosynthesis network, the purine metabolism network and the starch and sucrose metabolism network) also obtain F-measures with no neighbor information included, but they are less than 0.4.


Figure 5.7. F-measures obtained using the KStar classifier for the individual networks. Only networks that obtained an F-measure above 0.7 are shown.


Figure 5.8. F-measures obtained using the KStar classifier for the individual networks. Only networks that obtained an F-measure above 0.3 and below 0.7 are shown.


Figure 5.9. F-measures obtained using the KStar classifier for the individual networks. Only networks that obtained an F-measure above 0.1 and below 0.3 are shown.

With the networks that obtained the lower accuracies, the F-measures fluctuate depending on the amount of neighbor information that is included.

### 5.1.1.2 Results for Individual Chromosomes

When we look at each chromosome individually, we see that the results of classification vary from chromosome to chromosome, from the number of neighbors included, and from the classification algorithm being used.

### 5.1.1.2.1 Percent Correctly Classified

For all chromosomes, the percent of genomic features correctly classified to the correct network differed depending on the number of neighbors included in the dataset and the classification algorithm used.

When information about each genomic feature's neighbors is included, there is an improvement in the percent of features correctly classified for each of the classification algorithms for chromosome one (See Figure 5.10) The inclusion of information beyond each feature's twonearest neighbors has little effect on the percentage of correctly classified features. The MultiBoostAB algorithm using the J48 classifier obtains the highest percentage of correctly classified features.


Figure 5.10. Percent of genomic features classified for chromosome one using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome two, the KStar classifier appears to be influenced the most by the inclusion of neighbor information (See Figure 5.11). The Naïve Bayesian classifier does see an increase in its percentage of correctly classified features as more neighbors are included, but it is not as drastic as that observed with the KStar classifier. The percentage fluctuates using the MultiBoostAB algorithm with the J48 classifier. For the KStar classifier and the Naïve Bayesian classifier, the inclusion of neighbor information beyond each feature's nearest neighbors has little impact on the percentage of correctly classified features.

5.11. Percent of genomic features classified for chromosome two using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome three, the percent of genomic features correctly classified to the correct network increases when information about each genomic feature's nearest neighbors is included (See Figure 5.12). The MultiBoostAB using the J48 classifier sees the largest improvement. Increasing the number of neighbors beyond five, does not significantly improve the classification for the any of the classifiers.


Figure 5.12. Percent of genomic features classified for chromosome three using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

The inclusion of neighbor information for the genomic features of chromosome four increases the percentage of correctly classified genomic features for the KStar classifier and the Naïve Bayesian classifier, with the KStar classifier seeing the largest improvement (See Figure 5.13). The inclusion of neighbor information beyond each feature's nearest neighbors has little impact on these classifiers. Including neighbor information does not improve the percentage of correctly classified features for the MultiBoostAB algorithm using the J48 classifier.


Figure 5.13. Percent of genomic features classified for chromosome four using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome five, the inclusion of neighbor information is beneficial to the KStar classifier and the Naïve Bayesian classifier. The increase in the percentage of correctly classified features plateaus for these classifiers after the inclusion of information from each feature's nearest neighbors (See Figure 5.14). The MultiBoostAB algorithm using the J48 classifier does not see an increase in percentage of correctly classified features until information from each feature's two-nearest neighbors is included, then it obtains the higher percentages of correctly classified features.


Figure 5.14. Percent of genomic features classified for chromosome five using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome six, the inclusion of neighbor information increases the percentage of correctly classified features for all classifiers (See Figure 5.15). Inclusion of neighbor information beyond each feature's two-nearest neighbors has little impact on the percentages of correctly classified features.


Figure 5.15. Percent of genomic features classified for chromosome six using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome seven, the inclusion of neighbor information improves the percentage of correctly classified features for the KStar classifier and the Naïve Bayesian classifier. This increase is not observed for the MultiBoostAB algorithm using the J48 classifier until information from each feature's five-nearest neighbors is included (See Figure 5.16).


Figure 5.16. Percent of genomic features classified for chromosome seven using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome eight, all classifiers have an increase in the percentage of correctly classified features with the inclusion of neighbor information (See Figure 5.17). For the KStar classifier and the Naïve Bayesian classifier the inclusion of neighbor information beyond each feature's nearest neighbors has little impact on the percentages of correctly classified features.


Figure 5.17. Percent of genomic features classified for chromosome eight using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome nine, the inclusion of neighbor information improves the percentage of features correctly classified for both the KStar classifier and the MultiBoostAB algorithm using the J48 classifier (See Figure 5.18). The inclusion of neighbor information beyond each feature's nearest neighbors does not increase the percentage of correctly classified features for the KStar classifier. The inclusion of neighbor information beyond each feature's two-nearest neighbors has little impact on the MultiBoostAB algorithm using the J48 classifier. This classifier obtains the higher percentages at this point of information inclusion. There is not a drastic change in the percentage of correctly classified features using the Naïve Bayesian classifier as neighbor information is included.


Figure 5.18. Percent of genomic features classified for chromosome nine using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome ten, the inclusion of neighbor information is beneficial to the KStar classifier and the Naïve Bayesian classifier. The inclusion of information beyond each feature's nearest neighbors has little impact on the performance of the KStar classifier. For the MultiBoostAB algorithm using the J48 classifier, the inclusion of neighbor information generally causes a decrease in the percentage of correctly classified features. The exception occurs with the inclusion of information from each feature's five-nearest neighbors (See Figure 5.19).


Figure 5.19. Percent of genomic features classified for chromosome ten using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome eleven, the inclusion of neighbor information improves the percentage of correctly classified features for the KStar classifier and the Naïve Bayes classifier. The inclusion of information beyond each feature's nearest neighbors for the KStar classifier has little effect on the percentage of correctly classified features. For the Naïve Bayesian classifier, the inclusion of information beyond each feature's two-nearest neighbors has little effect on the percentage of correctly classified features. The MultiBoostAB algorithm using the J48 classifier, in general, sees a decrease in the percentage of correctly classified features with a spike seen when information from each feature's five-nearest neighbors is included. (See Figure 5.20).


Figure 5.20. Percent of genomic features classified for chromosome eleven using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome twelve, the KStar classifier and the Naïve Bayesian classifier observe an increase in the percentage of correctly classified features with the inclusion of neighbor information (See Figure 5.21). The inclusion of information beyond each feature's nearest neighbors for the KStar classifier has little effect on the percentage of correctly classified features. There is a decrease in the percentage of correctly classified features for the MultiBoostAB algorithm using the J 48 classifier when information from the number of neighbors increases from zero to two. As more neighbor information is included, an increase in the percentage of correctly classified features is seen for the MultiBoostAB algorithm using the J48 classifier.


Figure 5.21. Percent of genomic features classified for chromosome twelve using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome thirteen, the inclusion of more neighbor information improves the percentage of correctly classified features for the KStar and Naïve Bayesian classifiers (See Figure 5.22). The inclusion of information beyond each feature's nearest neighbors for the KStar classifier has little effect on the percentage of correctly classified features. For the MultiBoostAB algorithm using the J48 classifier, initially there is a decrease in the percentage of correctly classified features as the number of neighbors included goes from zero to two. When information from each feature's five-nearest neighbors is included, an increase in the percentage of correctly classified features is seen for the MultiBoostAB algorithm using the J48 classifier.


Figure 5.22. Percent of genomic features classified for chromosome thirteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome fourteen, for the KStar and Naïve Bayesian classifiers, there is an increase in the percentage of correctly classified genomic features as neighbor information is included (See Figure 5.23). The inclusion of information beyond each feature's nearest neighbors for the KStar classifier has little effect on the percentage of correctly classified features. The MultiBoostAB algorithm using the J 48 classifier sees a decrease in the percentage of correctly classified features followed by a steep increase when neighbor information is expanded from each feature's two- nearest neighbors to the ten-nearest neighbors.


Figure 5.23. Percent of genomic features classified for chromosome fourteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome fifteen, including neighbor information increases the percent of correctly classified genomic features for the KStar and the Naïve Bayesian classifiers (See Figure 5.24). The inclusion of information beyond each feature's nearest neighbors for the KStar classifier has little effect on the percentage of correctly classified features. The MultiBoostAB algorithm using the J48 classifier experiences a decrease in the percentage of correctly classified features when neighbor information is included.


Figure 5.24. Percent of genomic features classified for chromosome fifteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome sixteen, the addition of neighbor information does increase the percentage of correctly classified genomic features for the KStar classifier and the Naïve Bayesian classifier, with the KStar classifier seeing the larger increase (See Figure 5.25). The inclusion of information beyond each feature's nearest neighbors for the KStar classifier has little effect on the percentage of correctly classified features. The MultiBoostAB algorithm using the J48 classifier, in general, sees a decrease in the percentage of correctly classified features with a spike seen when information from each feature's five-nearest neighbors is included.


Figure 5.25. Percent of genomic features classified for chromosome sixteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

After analyzing all the chromosomes, the KStar classifier and the Naïve Bayesian classifier show similar patterns of performance for each chromosome. The MultiBoostAB algorithm using the J48 classifier does not show a consistent pattern of performance, but it did obtain the higher percentages of correctly classified features with the inclusion of more neighbor information. The KStar classifier obtained the higher percentages of correctly classified features with the inclusion of only each feature's nearest neighbors. Partitioning the classification process by chromosome
did not improve the percentage of correctly classified features. Each chromosome obtained close to a maximum of $45 \%$ using the KStar classifier, which is similar to the percentage that was obtained using the entire genome.

### 5.1.1.2.2. F-measures

To look at the accuracy of the classification process for each chromosome, the F-measures obtained for each chromosome can be analyzed.

For chromosome one, the F-measures are low. The higher F-measures are obtained with the MultiBoostAB algorithm using the J48 classifier, and the accuracy is affected by the number of neighbors included. The KStar classifier does not obtain F-measures above 0.11 regardless of the neighbor information included and the Naïve Bayesian classifier never achieves an F-measure above zero (See Figure 5.26).


Figure 5.26. F-measure for chromosome one using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome two, the KStar classifier achieves the higher F-measures. The inclusion of neighbor information improves its accuracy. The Naïve Bayesian classifier never achieves an F-
measure above zero, and the MultiBoostAB algorithm using the J 48 classifier only gets an Fmeasure that is not zero when information from the five-nearest neighbors is included (See Figure 5.27).


Figure 5.27. F-measure for chromosome two using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For the chromosome three, the F-measures are low. The MultiBoostAB algorithm using the J48 classifier obtains the higher F-measures with the inclusion of neighbor information having information from each feature's nearest neighbors obtaining the highest F-measure (See Figure 5.28). The F-measures for the KStar classifier improves when neighbor information is added, but the F-measures obtained are never above 0.1. The Naïve Bayesian classifier never achieves an F-measure above zero.


Figure 5.28. F-measure for chromosome three using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome four, the F-measures are low. The KStar classifier achieves the higher Fmeasures when neighbor information is included. The Naïve Bayesian classifier never obtains an F-measure above 0.15, but the inclusion of neighbor information from more than one neighbor is beneficial. The MultiBoostAB algorithm using the J48 classifier obtains its highest F-measure when no neighbor information is included (See Figure 5.29).


Figure 5.29. F-measure for chromosome four using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome five, the F-measures for the KStar classifier improve from zero to one when neighbor information is included. The F-measures obtained using the MultiBoostAB algorithm using the J48 classifier and the Naive Bayesian classifier improve to those levels when information from at least the two-nearest neighbors of each feature is included (See Figure 5.30).


Figure 5.30. F-measure for chromosome five using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome six, all of the classifiers observe an increase in F-measures as neighbor information is included (See Figure 5.31). For the KStar classifier and the MultiBoostAB algorithm using the J48 classifier, the increase observed goes from zero, when no neighbor information is included, to one regardless of how many neighbors are included. For the Naïve Bayesian classifier, information from more than one neighbor must be included to see the increase to one.


Figure 5.31. F-measure for chromosome six using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

The F-measures obtained using the different classifiers for chromosome seven are the lowest out of all the chromosomes (See Figure 5.32). The highest F-measure is obtained using the MultiBoostAB algorithm using the J48 classifier including information from each feature's tennearest neighbors.


Figure 5.32. F-measure for chromosome seven using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome eight, the KStar classifier obtains F-measures above 0.75 independent of the inclusion of neighbor information. The MultiBoostAB algorithm using the J48 classifier observes an increase in the F-measure when information from the two, five or ten-nearest neighbors of each feature is included (See Figure 5.33). The Naïve Bayesian classifier never obtains an Fmeasure above 0.1 independent of the inclusion of neighbor information.


Figure 5.33. F-measure for chromosome eight using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome nine, all the classifiers observe an increase in F-measures as more neighbor information is included. The KStar classifier and MultiBoostAB algorithm using the J48 classifier observe a steep increase with the inclusion of just one neighbor, with the F-measure for KStar classifier going from zero to one. The MultiBoostAB algorithm using the J48 does not obtain an F-measure until the inclusion of information from each feature's two-nearest neighbors. The F-measures obtained by the Naïve Bayesian classifier never exceed 0.4 (See Figure 5.34).


Figure 5.34. F-measure for chromosome nine using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome ten, the KStar classifier observes an increase in F-measures when neighbor information is included. The inclusion of neighbor information beyond each feature's nearest neighbors has little impact on the classifier. For the MultiBoostAB algorithm using the J48 classifier, the F-measure increases when information from the five-nearest neighbors is included. This increase is lost when information from the ten-nearest neighbors is included. The Naïve

Bayesian classifier observes little change in its F-measures as neighbor information is included (See Figure 5.35).


Figure 5.35. F-measure for chromosome ten using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

The F-measures for chromosome eleven have a similar pattern to those of chromosome ten (See Figure 5.36). The KStar classifier observes an increase in F-measures when neighbor information is included. For the MultiBoostAB algorithm using the J48 classifier, the F-measure increases when information from the five-nearest neighbors is included. This increase is lost when information from the ten-nearest neighbors is included. The F-measures for the Naïve Bayesian classifier remain zero independent of the inclusion of neighbor information.


Figure 5.36. F-measure for chromosome eleven using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome twelve, the F-measures obtained by the KStar classifier improve from near 0.7 to 1.0 with the inclusion of neighbor information. The inclusion of neighbor information beyond each feature's two-nearest neighbors has little impact on the KStar classifier. For the Naïve Bayesian classifier, the F-measures decrease as neighbor information is included. The Fmeasures obtained using the MultiBoostAB algorithm using the J48 classifier increase when information from each feature's five-nearest or the ten-nearest neighbors is included, with the higher F-measure obtained when including information from the ten-nearest neighbors (See Figure 5.37).


Figure 5.37. F-measure for chromosome twelve using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome thirteen, the F-measures obtained are low. The KStar classifier never obtains an F-measure above 0.1 (See Figure 5.38). The F-measures obtained using the Naïve Bayesian classifier increase as more neighbor information is included, but all remain below 0.16. The Fmeasures obtained using the MultiBoostAB algorithm using the J48 classifier increase with the inclusion of neighbor information, with the highest F-measure obtained when information from each feature's nearest neighbors is included.


Figure 5.38. F-measure for chromosome thirteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome fourteen, the KStar classifier observes an increase in the F-measure from zero to one as more neighbor information is included (See Figure 5.39). The Naïve Bayesian classifier never achieves an F-measure above zero regardless of the amount of neighbor information included. The F-measures obtained using the MultiBoostAB algorithm using the J48 classifier increase when information from each feature's five-nearest or the ten-nearest neighbors is included, with the higher F-measure obtained by including information from the ten-nearest neighbors (equal to one).


Figure 5.39. F-measure for chromosome fourteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome fifteen, the K-Star classifier observes an increase in the F-measure as neighbor information is included. The inclusion of neighbor information beyond each feature's nearest neighbors has little impact on the KStar classifier. The F-measures obtained by the other classifiers do not appear to be as affected by the inclusion of neighbor information. The Naïve Bayesian classifier experiences small increases when information each feature's nearest or twonearest neighbors is included (See Figure 5.40).


Figure 5.40. F-measure for chromosome fifteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome sixteen, the F-measures obtained for each classifier are very different (See Figure 5.41). Initially, as more neighbor information is included, the F-measure increases for the KStar classifier. The F-measure then drops when information from each feature's ten-nearest neighbors is included. For the MultiBoostAB algorithm, the F-measure increases when information from each feature's five-nearest neighbors is included then returns to zero when more neighbor information is included. The Naïve Bayesian classifier has the most consistent Fmeasures remaining zero or close to zero regardless of the inclusion of neighbor information.


Figure 5.41. F-measure for chromosome sixteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For a majority of the chromosome, the KStar classifier has the best accuracy, and the Naïve Bayesian classifier has the worst accuracy. Partitioning the classification process by chromosome does benefit the accuracy obtained by the classifiers.

### 5.1.2 Using Neighbor's Gene Ontology Term

A series of experiments was performed to determine if given the properties of chromosome number, start position, stop position, strand, neighbor's number, neighbor's Gene Ontology term, neighbor's strand, neighbor's distance (determined from the mid-point of each feature), and network involvement information from KEGG for genomic features (features being: genes, ARS, etc) for training could classification algorithms correctly classify network involvement. This was also performed including information from the nearest, two-, five- and ten-nearest neighbors of each feature. This was also repeated on the individual chromosomes. The mitochondrial chromosome is excluded. Having only one known network on the mitochondrial chromosome, classifiers cannot classify unary classes.

The results of the experiments are based on averages of ten different runs of each dataset being randomized and then split for training and testing. The training sets were set to contain $66 \%$ of the instances randomly selected from each dataset.

### 5.1.2.1 Results for the Entire Genome

When using only a genomic feature's chromosome number, start position, stop position, strand, and network involvement without neighbor information, we see that the percentage of genomic features correctly classified to the correct network varies depending on the classification algorithm used (See Figure 5.42). When no neighbor information is included, less than $20 \%$ of the genomic features are classified to the correct network regardless of the classifier used. As more neighbor information is added, the KStar classifier and the MultiBoostAB algorithm using
the J48 classifier increase their correctly classified percentage with the MultiBoostAB algorithm using the J48 classifier achieving the higher percentages. For both classifiers, the inclusion of neighbor information beyond each feature's two-nearest neighbors does not impact the percentage of correctly classified features. When the Naïve Bayesian classifier is used, the percentage of correctly classified genomic features drops as more neighbor information is included.


Figure 5.42. Percent of genomic features classified for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

When no neighbor information is included, all of the classifiers have poor agreement between the predicted and observed classification (See Figure 5.43). The KStar classifier and the MultiBoost AB algorithm using the J48 classifier achieve moderate agreements as more neighbor information is added with the MultiBoostAB algorithm using the J48 classifier achieving the higher Kappa statistic. The inclusion of neighbor information beyond each feature's two-nearest neighbors has little effect on the agreement between predicted and observed for the KStar
classifier and the MultiBoostAB algorithm using the J48 classifier. The Naïve Bayesian classifier never attains an agreement better than poor.


Figure 5.43. Kappa statistic for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

The inclusion of neighbor information affects the precision of the classification process. For the MultiBoostAB algorithm using the J48 classifier, the precision increases with the addition of neighbor information. The inclusion of neighbor information beyond each feature's two-nearest neighbors has little effect on the precision obtained by the MultiBoostAB algorithm using the J48 classifier. For the KStar classifier, the precision increases as more neighbor information is included, but it then decreases with the inclusion of each feature's ten-nearest neighbors. The Naïve Bayesian classifier never obtains high precision scores, but it does increase and plateau upon the inclusion of information from each feature's two-nearest neighbors (See Figure 5.44).


Figure 5.44. Precision for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

The recall scores obtained by the MultiBoostAB algorithm using the J 48 classifier improve with the inclusion of neighbor information. The inclusion of neighbor information beyond each feature's two-nearest neighbors has little effect on the recall obtained by the MultiBoostAB algorithm using the J48 classifier. The recall scores of the KStar classifier increases as more neighbor information is included then it decreases with the inclusion of each feature's ten-nearest neighbors. The recall scores obtained by the Naïve Bayes remain low regardless of the inclusion of neighbor information (See Figure 5.45).


Figure 5.45. Recall for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

F-measures obtained by the different classifiers mimic their precision and recall scores (See Figure 5.46). The accuracy of the classification of a genomic feature increases as more neighbor information is included for MultiBoostAB algorithm using the J48 classifier, with the inclusion of neighbor information beyond each feature's two-nearest neighbors having little effect on the Fmeasures obtained. The accuracy of the KStar classifier increases as more neighbor information is included then it decreases with the inclusion of each feature's ten-nearest neighbors. The accuracy of the Naïve Bayesian classifier is not greatly affected by the inclusion of neighbor information.


Figure 5.46. F-measure for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

When analyzing the area under the ROC curves, when no neighbor information is included, the MultiBoostAB algorithm using the J48 classifier is the only classifier that obtains an accuracy score that is not failing, but it is considered poor (See Figure 5.47). Upon the inclusion of neighbor information, the accuracy scores for both the KStar classifier and the MultiBoostAB algorithm using the J48 classifier improve to excellent. With the Naïve Bayesian classifier, the accuracy of the classification never improves above poor.


Figure 5.47. Area under ROC curve for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

In general, the MultiBoost AB algorithm using the J 48 classifier performs the best, obtaining the higher values in the metrics used.

To see the effect on the accuracy of classification for each network, the F-measures obtained by the MultiBoostAB algorithm using the J48 classifier for each network were analyzed. The table of the F-measures can be found in Appendix C. Most of the networks benefited from the addition of neighbor information (See Figures 5.48-5.50).

Three networks obtain F-measures when no information is included. These networks are the ribosome network, the polyunsaturated fatty acid biosynthesis network and the oxidative phosphorylation network. The F-measures are not above 0.5, so there is not a strong implication that feature location alone is an indicator for network involvement.


Figure 5.48. F-measures obtained using the MultiBoost AB algorithm using the J 48 classifier for the individual networks. Only networks that obtained an F-measure above 0.7 are shown.


Figure 5.49. F-measures obtained using the MultiBoostAB algorithm using the J48 classifier for the individual networks. Only networks that obtained an F-measure above 0.45 and below 0.7 are shown.


Figure 5.50. F-measures obtained using the MultiBoostAB algorithm using the J 48 classifier for the individual networks. Only networks that obtained an F-measure above 0.1 and below 0.45 are shown.

With the networks that obtained the lower accuracies, the F-measures fluctuate depending on the amount of neighbor information that is included.

### 5.1.2.2 Result for Individual Chromosomes

When we look at each chromosome individually, we see that the results of classification vary from chromosome to chromosome, from the number of neighbors included, and from the classification algorithm being used.

### 5.1.2.2.1 Percent Correctly Classified

For all chromosomes, the percent of genomic features correctly classified to the correct network differed depending on the number of neighbors included in the dataset and the classification algorithm used.

For chromosome one, all of the classifiers observe an increase in the percentage of correctly classified features when neighbor information is included (See Figure 5.51). The MultiBoostAB algorithm using the J48 classifier obtains the higher percentage of correctly classified features in all instances. For the MultiBoostAB algorithm using the J48 classifier, the inclusion of information beyond each feature's two-nearest neighbors has little impact on the percentage of correctly classified features. For the KStar classifier, the inclusion of information beyond each feature's nearest neighbors has little effect on the percentage of classified features. And, for the Naïve Bayesian classifier, the inclusion of information beyond each feature's five-nearest neighbors has little effect on the percentage of classified features.


Figure 5.51. Percent of genomic features classified for chromosome one using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome two, all of the classifiers observe an increase in the percentage of correctly classified features when neighbor information is included. The MultiBoostAB algorithm using the J48 classifier obtains the higher percentages (See Figure 5.52). For the MultiBoostAB algorithm using the J48 classifier, the inclusion of information beyond each feature's two-nearest neighbors has little impact on the percentage of correctly classified features. For the KStar classifier, the inclusion of information beyond each feature's nearest neighbors has little effect on the percentage of classified features. For the Naïve Bayesian classifier, the percentage of correctly classified features plateau after the inclusion of information from each feature's nearest neighbors; the percentage then increases upon the inclusion of information from each feature's ten-nearest neighbors.


Figure 5.52. Percent of genomic features classified for chromosome two using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome three, all of the classifiers observe an increase in the percentage of correctly classified features when neighbor information is included (See Figure 5.53). The MultiBoostAB algorithm using the J48 classifier obtains the highest percentages of correctly classified features with inclusion of neighbor information followed by the Naïve Bayesian classifier and the KStar classifier respectively.


Figure 5.53. Percent of genomic features classified for chromosome three using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome four, all of the classifiers observe an increase in the percentage of correctly classified features when neighbor information is included, with the MultiBoostAB algorithm using the J48 classifier obtaining the higher percentage (See Figure 5.54). For the MultiBoostAB algorithm using the J48 classifier, the inclusion of information beyond each feature's two-nearest neighbors has little impact on the percentages obtained. For the KStar classifier and the Naïve Bayesian classifier, inclusion of neighbor information beyond each feature's nearest neighbors has little impact on the percentages they obtain.


Figure 5.54. Percent of genomic features classified for chromosome four using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome five, all of the classifiers observe an increase in the percentage of correctly classified features when neighbor information is included with the MultiBoostAB algorithm using the J48 classifier obtaining the higher percentages (See Figure 5.55). For the MultiBoost AB algorithm using the J 48 classifier, the inclusion of information beyond each feature's five-nearest neighbors has little impact on the percentage of correctly classified features.

For the KStar classifier and the Naïve Bayesian classifier, the inclusion of information beyond each feature's nearest neighbors has little effect on the percentage of classified features.


Figure 5.55. Percent of genomic features classified for chromosome five using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome six, similar to other chromosomes, all of the classifiers observe an increase in the percentage of correctly classified features when neighbor information is included with the MultiBoostAB algorithm using the J48 classifier obtaining the higher percentages (See Figure 5.56). For the MultiBoostAB algorithm using the J48 classifier, the inclusion of information beyond each feature's two-nearest neighbors has little impact on the percentage of correctly classified features. For the KStar classifier and the Naïve Bayesian classifier, the inclusion of information beyond each feature's nearest neighbors has little effect on the percentage of classified features.


Figure 5.56. Percent of genomic features classified for chromosome six using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome seven, similar to other chromosomes, all of the classifiers observe an increase in the percentage of correctly classified features when neighbor information is included with the MultiBoostAB algorithm using the J48 classifier obtaining the higher percentages (See Figure 5.57). For the MultiBoostAB algorithm using the J48 classifier, the inclusion of information beyond each feature's two-nearest neighbors has little impact on the percentage of correctly classified features. For the KStar classifier and the Naïve Bayesian classifier, the inclusion of information beyond each feature's nearest neighbors has little effect on the percentage of classified features. The Naïve Bayesian classifier does not see a drastic improvement in the percentage of correctly classified features, and obtains the lower percentages.


Figure 5.57. Percent of genomic features classified for chromosome seven using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

The patterns of the percentage of correctly classified features for chromosome eight are similar to those seen in chromosome seven (See Figure 5.58). All of the classifiers observe an increase in the percentage of correctly classified features when neighbor information is included, with the MultiBoost AB algorithm using the J48 classifier obtaining the higher percentages. For the MultiBoost AB algorithm using the J 48 classifier, the inclusion of information beyond each feature's two-nearest neighbors has little impact on the percentage of correctly classified features. For the KStar classifier and the Naïve Bayesian classifier, the inclusion of information beyond each feature's nearest neighbors has little effect on the percentage of classified features. The Naïve Bayesian classifier does not see a drastic improvement in the percentage of correctly classified features, and obtains the lower percentages.


Figure 5.58. Percent of genomic features classified for chromosome eight using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome nine, all of the classifiers observe an increase in the percentage of correctly classified features when neighbor information is included. The MultiBoostAB algorithm using the J48 classifier obtains the higher percentages (See Figure 5.59). For the MultiBoostAB algorithm using the J48 classifier, the inclusion of information beyond each feature's two-nearest neighbors has little impact on the percentage of correctly classified features. For the KStar classifier, the inclusion of information beyond each feature's nearest neighbors has little effect on the percentage of classified features. For the Naïve Bayesian classifier, the percentage of correctly classified features plateau after the inclusion of information from each feature's two-nearest neighbors.


Figure 5.59. Percent of genomic features classified for chromosome nine using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome ten, the inclusion of neighbor information increases the percentage of correctly classified features for each classifier (See Figure 5.60). As more neighbor information is added, the percentage of correctly classified features increases for the MultiBoostAB algorithm using the J48 classifier. For the KStar classifier and the Naïve Bayesian classifier, the inclusion of information beyond each feature's nearest neighbors has little effect on the percentage of classified features. The Naïve Bayesian classifier does not see a drastic improvement in the percentage of correctly classified features, and obtains the lower percentages.


Figure 5.60. Percent of genomic features classified for chromosome ten using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

All of the classifiers experience increases in the percentage of correctly classified features when neighbor information is included for chromosome eleven (See Figure 5.61). For all classifiers, the inclusion of information beyond each feature's nearest neighbors has little impact on the percentage of correctly classified features obtained. The MultiBoostAB algorithm using the J48 classifier obtains the higher percentage of correctly classified features, and the Naïve Bayesian classifier obtains the lower.


Figure 5.61. Percent of genomic features classified for chromosome eleven using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome twelve, following trends established in other chromosomes, all of the classifiers observe an increase in the percentage of correctly classified features when neighbor information is included (See Figure 5.62). For the MultiBoostAB algorithm using the J48 classifier, the inclusion of information beyond each feature's five-nearest neighbors has little impact on the percentage of correctly classified features. For the KStar classifier and the Naïve Bayesian classifier, the inclusion of information beyond each feature's nearest neighbors has little effect on the percentage of correctly classified features. The Naïve Bayesian classifier does not see a drastic improvement in the percentage of correctly classified features, and obtains the lower percentages.


Figure 5.62. Percent of genomic features classified for chromosome twelve using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome thirteen, all of the classifiers observe an increase in the percentage of correctly classified features when neighbor information is included (See Figure 5.63). For all classifiers, the inclusion of information beyond each feature's nearest neighbors has little impact on the percentage of correctly classified features obtained. The MultiBoostAB algorithm using the J48
classifier obtains the higher percentage of correctly classified features, and the Naïve Bayesian classifier obtains the lower.


Figure 5.63. Percent of genomic features classified for chromosome thirteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome fourteen, the KStar classifier and the Naïve Bayesian classifier observe an increase in the percentage of correctly classified features upon the inclusion of information from each feature's nearest neighbor, then the percentages level off (See Figure 5.64). For the MultiBoostAB algorithm using the J48 classifier, an increase is not observed until information from each feature's two-or-more nearest neighbors is included, then it achieves the higher percentages of correctly classified features.


Figure 5.64. Percent of genomic features classified for chromosome fourteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

Similar to chromosome fourteen, the KStar classifier and the Naïve Bayesian classifier observe an increase in the percentage of correctly classified features upon the inclusion of information from each feature's nearest neighbor, then the percentages level off for chromosome fifteen (See Figure 5.65). For the MultiBoostAB algorithm using the J48 classifier, an increase is not observed until information from each feature's two-or-more nearest neighbors is included, then it achieves the higher percentages of correctly classified features.


Figure 5.65. Percent of genomic features classified for chromosome fifteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome sixteen, all of the classifiers observe an increase in the percentage of correctly classified features when neighbor information is included (See Figure 5.66). Similar to other chromosomes, for the MultiBoostAB algorithm using the J48 classifier, the inclusion of information beyond each feature's two-nearest neighbors has little impact on the percentage of correctly classified features. For the KStar classifier and the Naïve Bayesian classifier, the inclusion of information beyond each feature's nearest neighbors has little effect on the percentage of classified features. The Naïve Bayesian classifier does not see a drastic improvement in the percentage of correctly classified features, and obtains the lower percentages.


Figure 5.66. Percent of genomic features classified for chromosome sixteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

After analyzing all the chromosomes, all the classifiers benefitted from the inclusion of neighbor information. The MultiBoostAB algorithm using the J48 classifier performed better than the other classifiers, with the inclusion of information beyond each feature's two-nearest neighbors having little additional benefit to the process. The partitioning of the classification process by
chromosome did not affect it. Each chromosome obtained a percentage of correctly classified features that was similar to the percentage obtained using the entire genome.

### 5.1.2.2.2 F-measures

To look at the accuracy of the classification process for each chromosome, the F-measures obtained for each chromosome can be analyzed.

For chromosome one, the F-measures are low. The higher F-measures are obtained with the MultiBoost AB algorithm using the J48 classifier, and the accuracy increases with the addition of more neighbor information. The KStar classifier does not obtain F-measures above 0.1. The Fmeasures plateau for the KStar classifier after the inclusion of information from each feature's two-nearest neighbors. The Naïve Bayesian classifier never achieves an F-measure above zero (See Figure 5.67).


Figure 5.67. F-measure for chromosome one using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome two, the inclusion of neighbor information has an effect on all the classifiers. For the MultiBoostAB algorithm using the J48 classifier and the KStar classifier, this effect is seen upon the addition of information from each feature's nearest neighbors. At this point for the KStar classifier, the F-measures plateau. For the MultiBoostAB algorithm using the J48 classifier, the F-measures do not plateau until the addition of information from each feature's five-nearest neighbors. (See Figure 5.68). For the Naïve Bayesian classifier, an increase in the Fmeasure is not observed until information from each feature's ten-nearest neighbors is included.


Figure 5.68. F-measure for chromosome two using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome three, like chromosome one, the F-measures are low. The higher F-measures are obtained with the MultiBoostAB algorithm using the J48 classifier, and the accuracy is affected by the amount of neighbor information included. The F-measures level off after the inclusion of information from each feature's five-nearest neighbors. The KStar classifier does not obtain F-measures above 0.1. The inclusion of information beyond each feature's nearest neighbors has little impact on the F-measures obtained by the KStar classifier. The Naïve Bayesian classifier never achieves an F-measure above zero (See Figure 5.69).


Figure 5.69. F-measure for chromosome three using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome four, after information from each feature's two-nearest neighbors is included, the MultiBoostAB algorithm using the J48 classifier obtains higher F-measures, with its highest F-measure obtained when information from each feature's two-nearest neighbors is included. The KStar classifier also obtains its highest F-measure when information from each feature's two-nearest neighbors is included. The Naïve Bayesian classifier never obtains an F-measure above 0.1 (See Figure 5.70).


Figure 5.70. F-measure for chromosome four using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome five, all of the classifiers obtain drastic increases in their classification accuracy with the inclusion of neighbor information (Figure 5.71). The F-measures go from zero to one upon the inclusion of neighbor information for the KStar classifier and MultiBoostAB algorithm using the J48 classifier. After the inclusion of information from each feature's two-nearest neighbors, the Naïve Bayesian classifier obtains F-measures equal to one.


Figure 5.71. F-measure for chromosome five using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

Upon the addition of any neighbor information, all of the classifiers improve their F-measures from zero to one for chromosome six (See Figure 5.72.)


Figure 5.72. F-measure for chromosome six using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome seven, all the F-measures are low. The MultiBoostAB algorithm using the J48 classifier obtains the higher F-measures, but these values do not exceed 0.16 (See Figure 5.73). The KStar classifier never obtains F-measures above 0.04, and the F-measures remain zero for the Naïve Bayesian classifier.


Figure 5.73. F-measure for chromosome seven using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome eight, the inclusion of neighbor information improves the F-measures obtained by the KStar classifier and the MultiBoostAB algorithm using the J48 classifier. The KStar classifier obtains F-measures of one after the inclusion of information from each feature's nearest neighbors. This happens for the MultiBoostAB algorithm using the J48 classifier after the inclusion of information from each feature's two-nearest neighbors. The inclusion of neighbor information decreases the accuracy of the classification process for the Naïve Bayesian classifier (See Figure 5.74).


Figure 5.74. F-measure for chromosome eight using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome nine, the inclusion of neighbor information improves the F-measures obtained by all the classifiers. The KStar classifier obtains F-measures of one after the inclusion of information from each feature's nearest neighbors. This happens for the MultiBoostAB algorithm using the J48 classifier after the inclusion of information from each feature's twonearest neighbors. The improvement is not as drastic for the Naïve Bayesian classifier. Its F-
measures level off below 0.3 after the inclusion of information from each feature's two-nearest neighbors (See Figure 5.75).


Figure 5.75. F-measure for chromosome nine using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome ten, the inclusion of neighbor information improves the F-measures obtained by the KStar classifier and the MultiBoostAB algorithm using the J48 classifier, with the higher percentages obtained by the MultiBoostAB algorithm using the J48 classifier. For those classifiers, the inclusion of information from each feature's nearest neighbors has little effect on the F-measures obtained. For the Naïve Bayesian classifier, the inclusion of each feature's nearest neighbors improves its classification accuracy, but the addition of more than one neighbor lowers its accuracy (See Figure 5.76).


Figure 5.76. F-measure for chromosome ten using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome eleven, the inclusion of neighbor information improves the F-measures obtained by the KStar classifier and the MultiBoostAB algorithm using the J48 classifier with the higher percentages obtained by the MultiBoostAB algorithm using the J48 classifier. For the Naïve Bayesian classifier, there is a slight increase in F-measure with the inclusion of neighbor information, but it returns to zero with the inclusion of information from each feature's tennearest neighbors (See Figure 5.77).


Figure 5.77. F-measure for chromosome eleven using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome twelve, similar to chromosome eight, the inclusion of neighbor information improves the F-measures obtained by the KStar classifier and the MultiBoostAB algorithm using the J48 classifier. The KStar classifier obtains F-measures of one after the inclusion of information from each feature's nearest neighbors. This happens for the MultiBoostAB algorithm using the J48 classifier after the inclusion of information from each feature's twonearest neighbors. The inclusion of neighbor information decreases the accuracy of the classification process for the Naïve Bayesian classifier (See Figure 5.78).


Figure 5.78. F-measure for chromosome twelve using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome thirteen, the MultiBoostAB algorithm using the J48 classifier and the Naïve Bayesian classifier experience an increase in F-measures with the inclusion of neighbor information, with the higher F-measures being obtained by the MultiBoostAB algorithm using the J48 classifier. The inclusion of information beyond each feature's nearest neighbors has little effect on the accuracy of the MultiBoostAB algorithm using the J48 classifier and the Naïve Bayesian classifier. The KStar classifier never obtains an F-measure above zero (See Figure 5.79).


Figure 5.79. F-measure for chromosome thirteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome fourteen, the inclusion of neighbor information improves the F-measures obtained by the KStar classifier and the MultiBoostAB algorithm using the J48 classifier. The KStar classifier obtains F-measures of one after the inclusion of information from each feature's nearest neighbors. This happens for the MultiBoostAB algorithm using the J48 classifier after the inclusion of information from each feature's two-nearest neighbors. The accuracy of the Naïve Bayesian classifier does not improve above zero with the inclusion of neighbor information (See Figure 5.80).


Figure 5.80. F-measure for chromosome fourteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome fifteen, the inclusion of neighbor information improves the F-measures obtained by the KStar classifier and the MultiBoostAB algorithm using the J48 classifier, with the higher percentages obtained by the MultiBoostAB algorithm using the J48 classifier. For the KStar classifier, the inclusion of information from each feature's nearest neighbors has little effect on the F-measures obtained. For the MultiBoostAB algorithm using the J48 classifier, the inclusion of information from each feature's two-nearest neighbors has little effect on the Fmeasures obtained. The accuracy of the Naïve Bayesian classifier is not affected by the inclusion of neighbor information (See Figure 5.81).


Figure 5.81. F-measure for chromosome fifteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome sixteen, the KStar classifier observes an improvement in its classification accuracy with the inclusion of neighbor information with the inclusion of information beyond each feature's nearest neighbors having little effect on the F-measures obtained. The inclusion of information from the each feature's nearest, two nearest, or five-nearest neighbors benefits the accuracy of the MultiBoostAB algorithm using the J48 classifier. Upon the inclusion of
information from each feature's ten-nearest neighbors, the F-measure drops back to zero. The accuracy of the Naïve Bayesian classifier remains poor even with the inclusion of neighbor information (See Figure 5.82).


Figure 5.82. F-measure for chromosome sixteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

In general the MultiBoostAB algorithm using the J48 classifier obtained the higher F-measures for all the chromosomes. The partitioning of the classification process by chromosome benefited the accuracy of the classification for most of the classifiers. The classifiers obtained higher F-measures on individual chromosomes than what they were able to obtain using the entire genome. The decision trees that were generated can be found in Appendix H .

### 5.1.2.3 Removing Location Information

A series of experiments was performed to determine if given the properties of neighbor's number, neighbor's Gene Ontology term, neighbor's strand, neighbor's distance (determined from the mid-point of each feature), and network involvement information from KEGG for genomic features (features being: genes, ARS, etc) for training could the MultiBoostAB
algorithm using the J48 classifier could correctly classify network involvement. The MultiBoostAB algorithm using the J48 classifier was selected since it obtained the best performance in the prior section.

These experiments were performed to see if rules could be generated that would provided insight about a genomic feature's network involvement based only on what other features are near it regardless of that feature's location.

The rules generated did not incorporate the neighbor's Gene Ontology information. Neighbor distance was the most important attribute being selected for the root of the decision tree. Using information from each feature's two-nearest neighbors obtained the highest percentage of correctly classified features, but the percentage was low (See Figure 5.83).


Figure 5.83. The percentage of correctly classified features using the MultiBoostAB algorithm using the J48 classifier.

The accuracy of the classification was low regardless of the number of neighbors included (See Figure 5.84).


Figure 5.84. The F-measure using the MultiBoostAB algorithm using the J48 classifier.

### 5.1.3 Using Neighbor's Network Involvement

A series of experiments was performed to determine if given the genomic properties of chromosome number, start position, stop position, strand, neighbor's number, neighbor's strand, neighbor's distance (determined from the mid-point of each feature), neighbor's network involvement and the feature's network involvement information from KEGG for training could classification algorithms correctly classify network involvement. This was also performed including information from the nearest, two-, five- and ten- nearest neighbors of each feature. This was also repeated on the individual chromosomes. The mitochondrial chromosome is excluded. Having only one known network on the mitochondrial chromosome, classifiers cannot classify unary classes

The results of the experiments are based on averages of ten different runs of each dataset being randomized and then split for training and testing. The training sets were set to contain $66 \%$ of the instances randomly selected from each dataset.

### 5.1.3.1 Results for the Entire Genome

When using only a genomic feature's chromosome number, start position, stop position, strand, and network involvement without neighbor information, we see that the percentage of genomic features correctly classified to the correct net varies depending on the classification algorithm used (See Figure 5.85). When neighbor information is included, the percentage of correctly classified features increases for the KStar classifier. The percentage of correctly classified features does not change greatly with the inclusion of neighbor information for the MultiBoostAB algorithm using the J48 classifier until information from each feature's ten-nearest neighbors is included. At this point, it achieves the highest percentage of correctly classified features. For the Naïve Bayesian classifier, the percentage of correctly classified genomic features decreases when neighbor information is included.


Figure 5.85. Percent of genomic features classified for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

The agreement between the predicted and observed classification of the genomic features is affected by the inclusion of neighbor information. The KStar classifier achieves a moderate
agreement with the inclusion of neighbor information. The inclusion of neighbor information beyond each feature's nearest neighbors has little impact on the Kappa statistic for the KStar classifier. The MultiBoostAB algorithm using the J48 classifier attains an agreement that is moderate upon the addition of information from each feature's ten-nearest neighbors. The Naïve Bayesian classifier never attains an agreement above poor with the inclusion of neighbor information (See Figure 5.86).


Figure 5.86. Kappa statistic for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

The precision of feature classification is affected by the number of neighbors that is included (See Figure 5.87). For the KStar classifier, the inclusion of neighbor information improves the precision, with the inclusion of each feature's ten-nearest neighbors obtaining the highest precision. The MultiBoostAB algorithm using the J48 classifier obtains a precision score above 0.4 only with the inclusion of information from each feature's ten-nearest neighbors. The Naïve Bayesian classifier never obtains precision scores above 0.2 .


Figure 5.87. Precision for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

Recall is also affected by the number of neighbors that is included (See Figure 5.88). For the KStar classifier, the inclusion of neighbor information improves the recall, with the inclusion of each feature's ten-nearest neighbors obtaining the highest recall. The recall scores remain low for the other classifiers. The MultiBoostAB algorithm using the J48 classifier obtains a recall score above 0.4 after the inclusion of information from each feature's ten-nearest neighbors. The Naïve Bayesian classifier never obtains recall scores above 0.1; its higher recall score is obtained with the inclusion of information from each feature's two-nearest neighbors.


Figure 5.88. Recall for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

The F-measures of the classifiers mimic the precision and recall scores obtained by the classifiers (See Figure 5.89). For the KStar classifier, the inclusion of neighbor information improves the accuracy of the classification process, with the inclusion of each feature's ten-nearest neighbors obtaining the higher F-measure. The accuracy of the MultiBoostAB algorithm using the J48 classifier and is benefitted by the inclusion of neighbor information, with the inclusion of information from each feature's ten-nearest neighbors yielding the higher F-measures. For the Naïve Bayesian classifier, the higher recall score is obtained with the inclusion of information from each feature's two-nearest neighbors.


Figure 5.89. F-measure for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

When analyzing the area under the ROC curves, the accuracy improves to excellent with the inclusion of neighbor information for the KStar classifier. The inclusion of neighbor information beyond each feature's nearest neighbors has little impact on the area under the ROC curve for the KStar classifier. For the MultiBoostAB algorithm using the J48 classifier, the accuracy is poor with the inclusion of no neighbors. It improves to good with the inclusion of each feature's nearest neighbors and then it decreases with the inclusion of more neighbors, then increases to excellent upon the inclusion of information from each feature's ten-nearest neighbors. With the Naïve Bayesian classifier, the accuracy of the classification never improves above poor (See Figure 5.90).


Figure 5.90. Area under ROC curve for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

In general, the KStar classifier performs the best when trained using the genomic properties of chromosome number, start position, stop position, strand, neighbor's number, neighbor's strand, neighbor's distance (determined from the mid-point of each feature), neighbor's network involvement and the feature's network involvement information from KEGG.

To see the effect on the accuracy of classification for each network, the F-measures obtained by the classifier for each network were analyzed. The table of the F-measures can be found in Appendix D. Most of the networks benefited from the addition of neighbor information (See Figures 5.91-5.93).

Some networks attain F-measures that are above 0.6 with no neighbor information included implying that the location of these features alone is indicative of their network involvement. These networks are the ribosome network and the polyunsaturated fatty acid biosynthesis network. The inclusion of neighbor's network involvement for feature's of the ribosome network decreases the F-measures.

Other networks (the oxidative phosphorylation network, the SNARE interactions in vesicular transport network, the ubiquitin biosynthesis network, the proteasome network, the valine, leucine and isoleucine biosynthesis network, the glycerophospholipid metabolism network, the cell cycle network, the aminoacyl-tRNA biosynthesis network, the purine metabolism network and the starch and sucrose metabolism network) also obtain F-measures with no neighbor information included, but they are less than 0.4.

For networks that obtain lower F-measures, the F-measures fluctuate depending on how many neighbor's information is included.


Figure 5.91. F-measures obtained using the MultiBoostAB algorithm using the J 48 classifier for the individual networks. Only networks that obtained an F-measure above 0.7 are shown.


Figure 5.92. F-measures obtained using the MultiBoostAB algorithm using the J48 classifier for the individual networks. Only networks that obtained an F-measure above 0.3 and below 0.7 are shown.


Figure 5.93. F-measures obtained using the MultiBoostAB algorithm using the J48 classifier for the individual networks. Only networks that obtained an F-measure above 0.45 and below 0.7 are shown.

### 5.1.3.2 Result for Individual Chromosomes

When we look at each chromosome individually, we see that the results of classification vary from chromosome to chromosome, from the number of neighbors included, and from the classification algorithm being used.

### 5.1.3.2.1 Percent Correctly Classified

For all chromosomes, the percent of genomic features correctly classified to the correct network differed depending on the number of neighbors included in the dataset and the classification algorithm used.

For chromosome one, all of the classifiers observe an increase in the percentage of correctly classified features when neighbor information is included (See Figure 5.94). The MultiBoostAB algorithm using the J48 classifier obtains the higher percentage of correctly classified features. The percentages obtained by the MultiBoostAB algorithm using the J48 classifier plateau after information from each features two-nearest neighbors is included. The inclusion of neighbor information beyond each feature's nearest neighbors has little impact on the percentage of correctly classified features of the KStar classifier and the Naïve Bayesian classifier.


Figure 5.94. Percent of genomic features classified for chromosome one using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome two, all of the classifiers observe an increase in the percentage of correctly classified features when neighbor information is included with the MultiBoostAB algorithm using the J48 classifier obtained the larger percentages (See Figure 5.95). For the MultiBoostAB algorithm using the J48 classifier, the percentage of correctly classified features increases as more neighbor information is included. The inclusion of neighbor information beyond each feature's nearest neighbors has little impact on the percentage of correctly classified features of the KStar classifier and the Naïve Bayesian classifier.


Figure 5.95. Percent of genomic features classified for chromosome two using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome three, all of the classifiers observe an increase in the percentage of correctly classified features when neighbor information is included (See Figure 5.96). For the MultiBoostAB algorithm using the J48 classifier and the Naïve Bayesian classifier, the percentage of correctly classified features increases as more neighbor information is included. The inclusion of neighbor information beyond each feature's nearest neighbors has little impact on the percentage of correctly classified features of the KStar classifier.


Figure 5.96. Percent of genomic features classified for chromosome three using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome four, all of the classifiers observe an increase in the percentage of correctly classified features when neighbor information is included (See Figure 5.97). For the MultiBoostAB algorithm using the J48 classifier and the Naïve Bayesian classifier, the percentage of correctly classified features increases as more neighbor information is included. The MultiBoost AB algorithm using the J48 classifier obtains the higher percentages, and the Naïve Bayesian classifier obtains the lower percentages. The inclusion of neighbor information beyond each feature's nearest neighbors has little impact on the percentage of correctly classified features of the KStar classifier.


Figure 5.97. Percent of genomic features classified for chromosome four using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome five, all of the classifiers observe an increase in the percentage of correctly classified features when neighbor information is included (See Figure 5.98). For the MultiBoostAB algorithm using the J48 classifier, the percentage of correctly classified features increases as more neighbor information is included. The MultiBoostAB algorithm using the J48 classifier obtains the highest percentages. The inclusion of neighbor information beyond each feature's nearest neighbors has little impact on the percentage of correctly classified features of
the KStar classifier and the Naïve Bayesian classifier. The Naïve Bayesian classifier obtains the lowest percentages.


Figure 5.98. Percent of genomic features classified for chromosome five using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome six, all of the classifiers observe an increase in the percentage of correctly classified features when neighbor information is included, but the KStar classifier and the MultiBoost AB algorithm using the J48 classifier experience slight dips when information from each feature's two-nearest neighbors is included (See Figure 5.99). For the Naïve Bayesian classifier, the percentage of correctly classified features increases as more neighbor information is included. The MultiBoostAB algorithm using the J48 classifier obtains the higher percentages, and the Naïve Bayesian classifier obtains the lower percentages.


Figure 5.99. Percent of genomic features classified for chromosome six using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome seven, similar to other chromosomes, all of the classifiers observe an increase in the percentage of correctly classified features when neighbor information is included with the MultiBoost AB algorithm using the J48 classifier obtains the higher percentages (See Figure 5.100). The inclusion of neighbor information beyond each feature's nearest neighbors has little impact on the percentage of correctly classified features of the KStar classifier The Naïve Bayesian classifier does not see a drastic improvement in the percentage of correctly classified features, and it obtains the lower percentages.


Figure 5.100. Percent of genomic features classified for chromosome seven using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome eight, for all classifiers, the highest percentage of correctly classified features is obtained when information from each feature's nearest neighbors is included (See Figure 5.101). When more neighbor information is included, the percentage of correctly classified features decreases, but the percentages remain higher than when no neighbor information is included. The MultiBoostAB algorithm using the J48 classifier obtains the highest percentages, and the Naïve Bayesian classifier obtains the lowest percentages.


Figure 5.101. Percent of genomic features classified for chromosome eight using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome nine, the MultiBoostAB algorithm using the J48 classifier and the KStar classifier observe increases in the percentage of correctly classified features when neighbor information is included (See Figure 5.102). This increase is not seen in the Naïve Bayesian classifier until information from each feature's two-nearest neighbors has been included. The MultiBoostAB algorithm using the J48 classifier obtains the higher percentages after information from each feature's two-nearest neighbors is included. For the KStar classifier, the inclusion of information beyond each feature's nearest neighbors has little impact on the percentage of correctly classified features. The Naïve Bayesian classifier obtains the lower percentages.


Figure 5.102. Percent of genomic features classified for chromosome nine using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome ten, all of the classifiers observe an increase in the percentage of correctly classified features when neighbor information is included (See Figure 5.103). The MultiBoostAB algorithm using the J48 classifier obtains the higher percentages after information from each feature's two-nearest neighbors has been included. The Naïve Bayesian classifier obtains the lower percentages. For the KStar classifier, the inclusion of information beyond each feature's nearest neighbors has little impact on the percentage of correctly classified features.


Figure 5.103. Percent of genomic features classified for chromosome ten using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

All of the classifiers experience increases in the percentage of correctly classified features when neighbor information is included for chromosome eleven (See Figure 5.104). Similar to other chromosomes, the MultiBoostAB algorithm using the J48 classifier obtains the higher percentages, and the Naïve Bayesian classifier obtains the lower percentages. For the KStar classifier, the higher percentage of correctly classified features is obtained when information from each feature's nearest neighbors is included.


Figure 5.104. Percent of genomic features classified for chromosome eleven using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome twelve, the KStar classifier and the Naïve Bayesian classifier observe an increase in the percentage of correctly classified features when neighbor information is included (See Figure 5.105), with the inclusion of neighbor information beyond each feature's nearest neighbor having little impact on the percentage of correctly classified features. The MultiBoost AB algorithm using the J48 classifier experiences a dip in the percentage of correctly classified features when information from each feature's nearest neighbors is included, then it increases as more neighbor information is included, achieving the higher percentages.


Figure 5.105. Percent of genomic features classified for chromosome twelve using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome thirteen, all of the classifiers observe an increase in the percentage of correctly classified features when neighbor information is included (See Figure 5.106). Similar to other chromosomes, the MultiBoostAB algorithm using the J48 classifier obtains the higher percentages, and the Naïve Bayesian classifier obtains the lower percentages. For the MultiBoostAB algorithm using the J48 classifier, the percentage of correctly classified features increases as more neighbor information is included. The inclusion of neighbor information beyond each feature's nearest neighbors has little impact on the percentage of correctly classified features of the KStar classifier and the Naïve Bayesian classifier.


Figure 5.106. Percent of genomic features classified for chromosome thirteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome fourteen, all the classifiers obtain an increase in the percentage of correctly classified features with the inclusion of neighbor information (See Figure 5.107). For the MultiBoostAB algorithm using the J48 classifier and the Naïve Bayesian classifier, the percentage of correctly classified features increases as more neighbor information is included. The MultiBoostAB algorithm using the J 48 classifier obtains the higher percentages, and the Naïve Bayesian classifier obtains the lower percentages. The inclusion of neighbor information beyond each feature's nearest neighbors has little impact on the percentage of correctly classified features of the KStar classifier.


Figure 5.107. Percent of genomic features classified for chromosome fourteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome fifteen, all the classifiers obtain an increase in the percentage of correctly classified features with the inclusion of neighbor information with the MultiBoostAB algorithm using the J48 classifier obtains the higher percentage of correctly classified features (See Figure 5.108). For all classifiers, the highest percentages are obtained with the inclusion of information from each feature's ten-nearest neighbors.


Figure 5.108. Percent of genomic features classified for chromosome fifteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome sixteen, all of the classifiers observe an increase in the percentage of correctly classified features when neighbor information is included (See Figure 5.109) For the MultiBoost AB algorithm using the J48 classifier and the Naïve Bayesian classifier, the percentage of correctly classified features increases as more neighbor information is included. The MultiBoost AB algorithm using the J48 classifier obtains the highest percentages, and the Naïve Bayesian classifier obtains the lowest percentages. The inclusion of neighbor information beyond each feature's nearest neighbors has little impact on the percentage of correctly classified features of the KStar classifier.


Figure 5.109. Percent of genomic features classified for chromosome sixteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

After analyzing all the chromosomes, all the classifiers benefitted from the inclusion of neighbor information. The MultiBoostAB algorithm using the J48 classifier performed better than the other classifiers. This is different from what performed better using the entire genome. The partitioning of the classification process by chromosome benefitted the classification process. In general, for each chromosome, the MultiBoostAB algorithm using the J48 classifier was able to
obtain a higher percentage of correctly classified features than when compared to the results using the entire genome.

### 5.1.3.2.2 F-measures

To look at the accuracy of the classification process for each chromosome, the F-measures obtained for each chromosome can be analyzed.

For chromosome one, the F-measures are low. The F-measures obtained by both the KStar classifier and the MultiBoostAB algorithm using the J48 classifier fluctuate depending on how many neighbors are included. For both classifiers, the higher F-measures are obtained when information from each feature's ten-nearest neighbors is included. The Naïve Bayesian classifier never achieves an F-measure above zero (See Figure 5.110).


Figure 5.110. F-measure for chromosome one using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome two, F-measures above 0.5 are not obtained until information from each feature's ten-nearest neighbors is included for the MultiBoostAB algorithm using the J48
classifier and the KStar classifier (See Figure 5.111). The accuracy of the Naïve Bayesian classifier does not benefit from the inclusion of neighbor information.


Figure 5.111. F-measure for chromosome two using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome three, like chromosome one, the F-measures are low and they fluctuate for the MultiBoostAB algorithm using the J48 classifier and the KStar classifier. The higher F-measure for the MultiBoostAB algorithm using the J48 classifier is obtained with the inclusion of information from each feature's nearest neighbor. The higher F-measure for KStar classifier is obtained with the inclusion of information from each feature's five-nearest neighbors. The Naïve Bayesian classifier never achieves an F-measure above zero (See Figure 5.112).


Figure 5.112. F-measure for chromosome three using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome four, the inclusion of neighbor information benefits the accuracy of the KStar classifier and the MultiBoostAB algorithm using the J48 classifier. For both classifiers, the highest F-measures are obtained when information from each feature's five-nearest neighbors is included. The inclusion of neighbor information does not improve the accuracy of the Naïve Bayesian classifier (See Figure 5.113).


Figure 5.113. F-measure for chromosome four using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome five, the accuracy of all the classifiers improves drastically with the inclusion of information from each feature's ten-nearest neighbors (Figure 5.114). The F-measures go from zero to one.


Figure 5.114. F-measure for chromosome five using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

Similar to chromosome five, upon the addition of information from each feature's ten-nearest neighbors, all of the classifiers improve their F-measures to one for chromosome six (See Figure 5.115).


Figure 5.115. F-measure for chromosome six using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome seven, the inclusion of some neighbor information is beneficial for the KStar classifier and the MultiBoostAB algorithm using the J48 classifier. Both classifiers experience decreases in the F-measures with the inclusion of information from each feature's ten-nearest neighbors (See Figure 5.116). The Naïve Bayesian classifier never obtains an F-measure above zero.


Figure 5.116. F-measure for chromosome seven using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome eight, the inclusion of neighbor information improves the F-measures obtained by the KStar classifier and the MultiBoostAB algorithm using the J48 classifier. For both classifiers, the inclusion of information beyond each feature's nearest neighbors has little impact on the F-measures obtained. The F-measures obtained by the Naïve Bayesian classifier fluctuate depending on the amount of neighbor information that is included with the highest Fmeasure being obtained with the inclusion of information from each feature's nearest neighbors (See Figure 5.117).


Figure 5.117. F-measure for chromosome eight using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome nine, upon the inclusion of information from each feature's ten-nearest neighbors, the KStar classifier and the MultiBoostAB algorithm using the J48 classifier experience a drastic increase in their F-measures. The increase obtained by the Naïve Bayesian classifier is not as great (See Figure 5.118).


Figure 5.118. F-measure for chromosome nine using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome ten, the inclusion of neighbor information improves the F-measures obtained by the KStar classifier and the MultiBoostAB algorithm using the J48 classifier (both achieve Fmeasures of 1.0 ), but the F-measures with the inclusion of each feature's ten-nearest neighbors decrease. The Naïve Bayesian classifier never obtains an F-measure above zero (See Figure 5.119).


Figure 5.119. F-measure for chromosome ten using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome eleven, as more neighbor information is included, the F-measures for the MultiBoost AB algorithm using the J48 classifier increases. The F-measures obtained using the KStar classifier fluctuate depending on the number of neighbors included. The Naïve Bayesian classifier never obtains an F-measure above zero (See Figure 5.120).


Figure 5.120. F-measure for chromosome eleven using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome twelve, the inclusion of neighbor information decreases the F-measures obtained by the KStar classifier until information from each feature's ten-nearest neighbors is included. The inclusion of information from each feature's nearest neighbors or the two-nearest neighbors decreases the F-measures obtained using the MultiBoostAB algorithm using the J48 classifier. With the inclusion of more neighbor information, the F-measures then rise for the MultiBoostAB algorithm using the J48 classifier. The F-measures obtained using the Naïve Bayesian classifier with the inclusion of neighbor information are not as high as that obtained when no neighbor information is included (See Figure 5.121).


Figure 5.121. F-measure for chromosome twelve using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome thirteen, the F-measures are low. For the MultiBoostAB algorithm using the J48 classifier and the KStar classifier, the higher F-measures are obtained when information from each feature's five-nearest neighbors is included. For the Naïve Bayesian classifier, the Fmeasures increase as information from more than two-nearest neighbors is included (See Figure 5.122).


Figure 5.122. F-measure for chromosome thirteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome fourteen, upon the addition of information from each feature's ten-nearest neighbors, the KStar classifier and the MultiBoostAB algorithm using the J48 classifier drastically improve their F-measures, going from near 0.1 to 1.0. The F-measures for the Naïve Bayesian classifier fluctuate, obtaining the highest F-measure with the inclusion of information from each feature's nearest neighbors (See Figure 5.123).


Figure 5.123. F-measure for chromosome fourteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome fifteen, the addition of neighbor information improves the accuracy of both the KStar classifier and the MultiBoostAB algorithm using the J48 classifier. The MultiBoostAB algorithm using the J48 classifier obtains the higher F-measures. The accuracy of the Naïve Bayesian classifier is improved slightly with only the inclusion of information from each feature's two-nearest neighbors (See Figure 5.124).


Figure 5.124. F-measure for chromosome fifteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome sixteen, the F-measures are low. The KStar classifier never obtains an Fmeasure above zero. With the addition of information from each feature's ten-nearest neighbors, the MultiBoostAB algorithm using the J48 classifier improves its F-measure from zero to above 0.1. The Naïve Bayesian classifier obtains its highest F-measure with the inclusion of each feature's five-nearest neighbors (See Figure 5.125).


Figure 5.125. F-measure for chromosome sixteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

From this analysis the MultiBoostAB algorithm obtains the higher accuracy in the classification process when neighbor information is included for most of the chromosomes. The partitioning of the classification process by chromosome produced in some cases higher F-measures than those that were obtained using the entire genome.

### 5.2 Classifying Genomic Features to Gene Ontology Terms

The results of experiments that determined if altering the type of neighbor information given the classifiers altered the accuracy of the classification process follow.

### 5.2.1 Using Neighbor's SGDID

A series of experiments was performed to determine if given the properties of chromosome number, start position, stop position, strand, neighbor number, neighbor's SGDID, neighbor's strand, neighbor's distance (determined from the mid-point of each feature), Gene Ontology aspect and Gene Ontology term for genomic features (features being: genes, ARS, etc) could classification algorithms correctly classify the Gene Ontology term. This was also performed including information from the nearest, two-, five- and ten-nearest neighbors of each feature. This was also repeated on the individual chromosomes.

The results of the experiments are based on averages of ten different runs of each dataset being randomized and then split for training and testing. The training sets were set to contain $66 \%$ of the instances randomly selected from each dataset.

### 5.2.1.1 Results for Entire Genome

None of the classifiers obtained a percentage of correctly classified features that was above $30 \%$ (See Figure 5.126). As more neighbor information is added, the KStar classifier and the MultiBoost AB algorithm using the J 48 classifier increase their correctly classified percentage,
with the MultiBoostAB algorithm using the J48 classifier achieving the higher percentages. For the MultiBoostAB algorithm using the J48 classifier, the inclusion of neighbor information beyond each feature's nearest neighbors does not impact the percentage of correctly classified features. When the Naïve Bayesian classifier is used, the percentage of correctly classified genomic features is not affected by the addition of neighbor information.


Figure 5.126. Percent of genomic features classified for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

Regardless of the inclusion of neighbor information, none of the classifiers obtain a Kappa statistic above 0.25 . This indicates that under all of the tested conditions, all of the classifiers obtained poor agreement between predicted and observed classification (See Figure 5.127).


Figure 5.127. Kappa statistic for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

The precision scores are low under all test conditions for all of the classifiers. For the KStar classifier, the inclusion of neighbor information improves the classification precision. For the MultiBoostAB algorithm using the J48 classifier, the inclusion of neighbor information decreases the precision of the classification. The precision scores obtained by the Naïve Bayesian classifier are zero (See Figure 5.128).


Figure 5.128. Precision for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For all classifiers, the recall scores are low regardless of the inclusion of neighbor information. The recall scores of the KStar classifier increases as more neighbor information is included but they remain low. The recall scores obtained by the MultiBoostAB algorithm using the J48 classifier increase with the inclusion neighbor information. For the Naïve Bayesian classifier, the recall scores are zero (See Figure 5.129).


Figure 5.129. Recall scores for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

F-measures obtained by the different classifiers mimic their precision scores (See Figure 5.130). The F-measures are low under all test conditions for all of the classifiers. For the KStar classifier, the inclusion of neighbor information improves the accuracy of classification precision. For the MultiBoostAB algorithm using the J48 classifier, the inclusion of neighbor information decreases the accuracy of the classification. The F-measures obtained by the Naïve Bayesian classifier are zero.


Figure 5.130. F-measure for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

When analyzing the area under the ROC curves, when no neighbor information is included, the Naïve Bayesian classifier is the only classifier that obtains an accuracy score that is deemed good. This result does not change with the inclusion of neighbor information. The KStar classifier improves from poor to excellent with the inclusion of neighbor information, and it obtains the higher accuracy scores. For the MultiBoostAB algorithm using the J48 classifier, the accuracy improves from poor to good with the inclusion of neighbor information (See Figure 5.131).


Figure 5.131. Area under ROC curve for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

None of the classifiers performed well based on the metrics used. Since the MultiBoostAB algorithm using the J 48 classifier obtained the higher percentage of correctly classified features, it was used in the analysis of the classification of features to individual Gene Ontology terms.

To see the effect on the accuracy of classification for each network, the F-measures obtained by the MultiBoostAB algorithm using the J48 classifier for each network were analyzed. The table of the F-measures can be found in Appendix D. Most of the Gene Ontology term groups did not benefit from the addition of neighbor information (See Figures 5.132 and 5.133). Only three Gene Ontology term groups (biological process, cytoplasm and molecular function) showed an increase in the classification of features to the groups with the addition of neighbor information.


Figure 5.132. F-measures obtained using the MultiBoostAB algorithm using the J 48 classifier for the individual Gene Ontology (GO) terms.


Figure 5.133. F-measures obtained using the MultiBoostAB algorithm using the J48 classifier for the individual Gene Ontology (GO) terms.

### 5.2.1.2 Result for Individual Chromosomes

When we look at each chromosome individually, we see that the results of classification vary from chromosome to chromosome, from the number of neighbors included, and from the classification algorithm being used.

### 5.2.1.2.1 Percent Correctly Classified

For all chromosomes, the percent of genomic features correctly classified to the correct Gene Ontology term did not differ drastically as neighbor information is included, but the classifiers did perform differently.

For chromosome one, the inclusion of neighbor information increases the percentage of correctly classified features for all classifiers (See Figure 5.134). For the MultiBoostAB algorithm using the J48 classifier, the inclusion of neighbor information from each feature's ten-nearest neighbors drastically increases the percentage of correctly classified features. For the KStar classifier and the Naïve Bayesian classifier, the inclusion of neighbor information beyond each feature's nearest neighbor has little impact on the classification process. The KStar classifier obtains the lower percentages of correctly classified features


Figure 5.134. Percent of genomic features classified for chromosome one using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome two, the MultiBoostAB algorithm using the J48 classifier and the KStar classifiers are the most affected by the inclusion of neighbor information. The inclusion of each feature's nearest and two-nearest neighbors produces the higher percentages of correctly classified features. The KStar classifier obtains the lowest percentages of correctly classified features, and the inclusion of information beyond each feature's nearest neighbors has little impact. For the Naïve Bayesian classifier, the percentage of correctly classified features is no affected by the inclusion of neighbor information (See Figure 5.135).

5.135. Percent of genomic features classified for chromosome two using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome three, the inclusion of neighbor information increases the percentage of correctly classified features for the KStar classifier and the Naïve Bayesian classifier. The inclusion of neighbor information beyond each feature's nearest neighbors has little effect on the percentage of correctly classified feature by the Naïve Bayesian classifier. The percentage of correctly classified features for the MultiBoostAB algorithm using the J48 classifier is not as affected by the inclusion of neighbor information (See Figure 5.136), and it obtains the highest percentages. The KStar classifier obtains the lower percentages.


Figure 5.136. Percent of genomic features classified for chromosome three using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome four, the MultiBoostAB algorithm using the J48 classifier and the KStar classifiers are the most affected by the inclusion of neighbor information. The inclusion of each feature's nearest and two-nearest neighbors produces the higher percentages of correctly classified features. The KStar classifier obtains the lower percentages of correctly classified features, and the inclusion of information beyond each feature's nearest neighbors has little impact. For the Naïve Bayesian classifier, the percentage of correctly classified features decreases slightly with the inclusion of neighbor information (See Figure 5.137).


Figure 5.137. Percent of genomic features classified for chromosome four using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome five, the percentage of correctly classified features increases for all classifiers with the inclusion of neighbor information. The MultiBoostAB algorithm using the J48 classifier obtains the higher percentages, with the inclusion of each feature's nearest and two-nearest neighbors producing the higher percentages of correctly classified features. The KStar classifier obtains the lower percentages of correctly classified features. The inclusion of neighbor information beyond each feature's nearest neighbors has little impact on the KStar classifier and the Naïve Bayesian classifier (See Figure 5.138).


Figure 5.138. Percent of genomic features classified for chromosome five using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome six, the inclusion of neighbor information affects the Naïve Bayesian classifier the most. The inclusion of neighbor information beyond each feature's nearest neighbors has little impact on the percentages obtained by the Naïve Bayesian classifier. The KStar classifier obtains a slight increase in the percentage of correctly classified features as more neighbor information is included. The MultiBoostAB algorithm using the J 48 classifier obtains the higher percentages of correctly classified features, but the percentages are not affected by the inclusion of neighbor information until information from each feature's ten-nearest neighbors is included, then, it decreases (See Figure 5.139).


Figure 5.139. Percent of genomic features classified for chromosome six using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome seven, the MultiBoostAB algorithm using the J48 classifier and the KStar classifiers are the most affected by the inclusion of neighbor information. The inclusion of each feature's nearest and two-nearest neighbors produces the higher percentages of correctly classified features. The KStar classifier obtains the lower percentages of correctly classified features, and the inclusion of information beyond each feature's nearest neighbors has little impact. For the Naïve Bayesian classifier, the percentage of correctly classified features decreases slightly with the inclusion of neighbor information (See Figure 5.140).


Figure 5.140. Percent of genomic features classified for chromosome seven using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome eight, the KStar classifier is affected most by the inclusion of neighbor information, with the inclusion of neighbor information increasing the percentage of correctly classified features. The MultiBoostAB algorithm using the J48 classifier obtains the higher percentages of correctly classified features. The MultiBoostAB algorithm using the J48 classifier and the KStar classifier do not experience drastic changes in the percentage of correctly classified features with the inclusion information (See Figure 5.141).


Figure 5.141. Percent of genomic features classified for chromosome eight using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome nine, the MultiBoostAB algorithm using the J48 classifier achieves the higher percentages of correctly classified features compared the other classifiers. The percentage of correctly classified features remains constant for the MultiBoostAB algorithm using the J48 classifier until the inclusion of neighbor information from each feature's five-nearest neighbors, then, the percentages drop. The inclusion of neighbor information slightly increase the percentage of correctly classified features for the KStar classifier and the Naïve Bayesian classifier (See Figure 5.142).


Figure 5.142. Percent of genomic features classified for chromosome nine using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome ten, the percentage of correctly classified features increases for all classifiers with the inclusion of neighbor information. The MultiBoostAB algorithm using the J48 classifier obtains the higher percentages, with the inclusion of each feature's nearest and two-nearest neighbors producing the higher percentages of correctly classified features. The KStar classifier obtains the lower percentages of correctly classified features. The inclusion of neighbor information beyond each feature's nearest neighbors has little impact on the KStar classifier and the Naïve Bayesian classifier (See Figure 5.143).


Figure 5.143. Percent of genomic features classified for chromosome ten using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome eleven, the percentage of correctly classified features increases for the MultiBoostAB algorithm using the J48 classifier and the KStar classifier with the inclusion of neighbor information. The MultiBoostAB algorithm using the J48 classifier obtains the higher percentages, with the inclusion of each feature's nearest and two-nearest neighbors producing the higher percentages of correctly classified features, but this increase is lost as more neighbor information is included. The KStar classifier obtains the lower percentages of correctly classified features. The inclusion of neighbor information beyond each feature's nearest neighbors has little impact on the KStar classifier. The percentage of correctly classified features for the Naïve Bayesian classifier is not affected by the inclusion of neighbor information (See Figure 5.144).


Figure 5.144. Percent of genomic features classified for chromosome eleven using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome twelve, the MultiBoostAB algorithm using the J48 classifier and the KStar classifiers are the most affected by the inclusion of neighbor information. The inclusion of each feature's nearest and two-nearest neighbors produces the higher percentages of correctly classified features. The KStar classifier obtains the lower percentages of correctly classified features, and the inclusion of information beyond each feature's nearest neighbors has little impact. For the Naïve Bayesian classifier, the percentage of correctly classified features is not affected by the inclusion of neighbor information (See Figure 5.145).


Figure 5.145. Percent of genomic features classified for chromosome twelve using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome thirteen, the MultiBoostAB algorithm using the J48 classifier and the KStar classifiers are the most affected by the inclusion of neighbor information. The inclusion of each feature's nearest and two-nearest neighbors produces the higher percentages of correctly classified features. The KStar classifier obtains the lower percentages of correctly classified features, and the inclusion of information beyond each feature's nearest neighbors has little impact. For the Naïve Bayesian classifier, the percentage of correctly classified features is not affected by the inclusion of neighbor information (See Figure 5.146).


Figure 5.146. Percent of genomic features classified for chromosome thirteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome fourteen, for the KStar classifier, there is an increase in the percentage of correctly classified features with the inclusion of neighbor information, with the inclusion of neighbor information beyond each feature's nearest neighbors having little impact. The higher percentages are obtained by the MultiBoostAB algorithm using the J48 classifier. The inclusion of information from each feature's ten-nearest neighbors decreases the percentage of correctly classified features for the MultiBoostAB algorithm using the J 48 classifier and the Naïve Bayesian classifier (See Figure 5.147).


Figure 5.147. Percent of genomic features classified for chromosome fourteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome fifteen, similar to other chromosomes, the MultiBoostAB algorithm using the J48 classifier and the KStar classifiers are the most affected by the inclusion of neighbor information. The inclusion of each feature's nearest and two-nearest neighbors produces the higher percentages of correctly classified features. The KStar classifier obtains the lower percentages of correctly classified features, and the inclusion of information beyond each feature's nearest neighbors has little impact. For the Naïve Bayesian classifier, the percentage of correctly classified features is not affected by the inclusion of neighbor information (See Figure 5.148).


Figure 5.148. Percent of genomic features classified for chromosome fifteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

Like other chromosomes, for chromosome sixteen, the MultiBoostAB algorithm using the J48 classifier obtains the higher percentages of correctly classified features. The inclusion of information from each feature's nearest and two-nearest neighbors increases the percentage of features correctly classified, but this increase is lost as more neighbor information is included. For the KStar classifier, the inclusion of neighbor information increases the percentage of correctly classified features. For the Naïve Bayesian classifier, the percentage of correctly classified features is not affected by the inclusion of neighbor information (See Figure 5.149).


Figure 5.149. Percent of genomic features classified for chromosome sixteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For the mitochondrial chromosome, the MultiBoostAB algorithm using the J48 classifier obtains an increase in the percentage of correctly classified features with the inclusion of information from two or more nearest neighbors of each feature. The inclusion of neighbor information does not have a drastic affect on the KStar classifier and the Naïve Bayesian classifier. The KStar classifier obtains the lower percentages of correctly classified features (See Figure 5.150).


Figure 5.150. Percent of genomic features classified for the mitochondrial chromosome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For most of the chromosomes, the MultiBoostAB algorithm using the J48 classifier obtained a higher percentage of correctly classified features followed by the Naïve Bayesian classifier then the KStar classifier. The mitochondrial chromosome is the only chromosome that has classifiers that obtain percentages of correctly classified features above $60 \%$. The partitioning of the classification process by chromosome, in general, did not improve the percentages of correctly classified features for the classifiers.

### 5.2.1.2.2. F-measures

To look at the accuracy of the classification process for each chromosome, the F-measures obtained for each chromosome can be analyzed. The accuracy of the classification process is affected by the inclusion of neighbor information, and the classifier is used. The F-measures for the chromosomes are low.

For chromosome one, the highest F-measure is obtained by the MultiBoostAB algorithm using the J48 classifier with the inclusion of information from each feature's nearest neighbor. The inclusion of neighbor information increases the F-measures obtained by the Naïve Bayesian classifier. The KStar classifier obtains the lower F-measures, and it obtains slightly increased Fmeasures with the inclusion of neighbor information (See Figure 5.151).


Figure 5.151. F-measure for chromosome one using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome two, the highest F-measure is obtained by the Naïve Bayesian classifier with information from each feature's five-nearest neighbors is included. The F-measures for the MultiBoostAB algorithm using the J48 classifier fluctuate depending on the amount of neighbor information that is included. The largest decrease in the F-measures is seen when information from each feature's two-nearest neighbors is included for the MultiBoostAB algorithm using the J48 classifier. The inclusion of neighbor information increases the F-measures for the KStar classifier (See Figure 5.152).


Figure 5.152. F-measure for chromosome two using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome three, the inclusion of neighbor information increases the F-measures obtained by the Naïve Bayesian classifier. The higher F-measure is obtained with the Naïve Bayesian classifier with the inclusion of information from each feature's five-nearest neighbors. The Fmeasures for the MultiBoostAB algorithm using the J48 classifier fluctuate depending on the amount of neighbor information that is included. The KStar classifier has the lower F-measures. The inclusion of information from each feature's five-or-more nearest neighbors increases the Fmeasures obtained by the KStar classifier (See Figure 5.153).


Figure 5.153. F-measure for chromosome three using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome four, the inclusion of neighbor information increases the F-measures obtained by the Naïve Bayesian classifier and the KStar classifier. This increase is not seen for the MultiBoostAB algorithm using the J48 classifier until information from each feature's fivenearest or ten-nearest neighbors is included (See Figure 5.154).


Figure 5.154. F-measure for chromosome four using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome five, the inclusion of neighbor information improves the F-measures for the Naïve Bayesian classifier and the KStar classifier, with the higher F-measures obtained by the Naïve Bayesian classifier. The inclusion of neighbor information from each feature's nearest and two-nearest neighbors decreases the F-measures for the MultiBoostAB algorithm using the J48 classifier, but the F-measures then rebound upon the inclusion of information from each feature's five-or-more-nearest neighbors (See Figure 5.155).


Figure 5.155. F-measure for chromosome five using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome six, the inclusion of neighbor information increased the F-measures obtained by all the classifiers. The higher F-measures are obtained by the MultiBoostAB algorithm using the J48 classifier. The inclusion of neighbor information beyond each feature's five-nearest neighbors has little impact on the accuracy of classification process for the MultiBoostAB algorithm using the J48 classifier. The inclusion of neighbor information from each feature's nearest neighbors has little impact the accuracy of the classification process for the KStar classifier and the Naïve Bayesian classifier (See Figure 5.156).


Figure 5.156. F-measure for chromosome six using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome seven, like chromosome four, the inclusion of neighbor information increases the F-measures obtained by the Naïve Bayesian classifier and the KStar classifier. This increase is not seen for the MultiBoostAB algorithm using the J48 classifier until information from each feature's five-nearest or ten-nearest neighbors is included (See Figure 5.157).


Figure 5.157. F-measure for chromosome seven using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome eight, the inclusion of neighbor information increases the F-measures obtained by the Naïve Bayesian classifier and the KStar classifier. The inclusion of neighbor information decreases the F-measures obtained by the MultiBoostAB algorithm using the J48 classifier, with the largest decrease observed upon the inclusion of neighbor information from each feature's nearest neighbors (See Figure 5.158).


Figure 5.158. F-measure for chromosome eight using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

Similar to other chromosomes, for chromosome nine, the inclusion of neighbor information increases the F-measures obtained by the Naïve Bayesian classifier and the KStar classifier. This increase is not seen for the MultiBoostAB algorithm using the J48 classifier until information from each feature's five-nearest or ten-nearest neighbors is included (See Figure 5.159).


Figure 5.159. F-measure for chromosome nine using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome ten, the inclusion of neighbor information increases the F-measures obtained by the Naïve Bayesian classifier and the KStar classifier. The inclusion of information from each feature's nearest and two-nearest neighbors decreases the F-measures obtained by the MultiBoostAB algorithm using the J48 classifier. With the inclusion of information from the five- or ten-nearest neighbors of each feature, the F-measures then increase for the MultiBoostAB algorithm using the J48 classifier (See Figure 5.160).


Figure 5.160. F-measure for chromosome ten using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

Like other chromosomes, the inclusion of neighbor information increases the F-measures obtained by the Naïve Bayesian classifier and the KStar classifier for chromosome eleven. This increase is not seen for the MultiBoostAB algorithm using the J 48 classifier until information from each feature's five-nearest or ten-nearest neighbors is included (See Figure 5.161).


Figure 5.161. F-measure for chromosome eleven using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome twelve, the inclusion of neighbor information increases the F-measures obtained by the Naïve Bayesian classifier and the KStar classifier. The inclusion of information from each feature's nearest and two-nearest neighbors decreases the F-measures obtained by the MultiBoostAB algorithm using the J48 classifier. With the inclusion of information from the five- or ten-nearest neighbors of each feature, the F-measures then increase for the MultiBoostAB algorithm using the J48 classifier (See Figure 5.162).


Figure 5.162. F-measure for chromosome twelve using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome thirteen, the inclusion of neighbor information increases the F-measures obtained by the Naïve Bayesian classifier and the KStar classifier. The inclusion of information from each feature's nearest and two-nearest neighbors decreases the F-measures obtained by the MultiBoostAB algorithm using the J48 classifier, with the inclusion of information from the fiveor ten-nearest neighbors of each feature, the F-measures then increase and are the highest Fmeasures obtained by any classifier (See Figure 5.163).


Figure 5.163. F-measure for chromosome thirteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome fourteen, the inclusion of neighbor information increases the F-measures obtained by the Naïve Bayesian classifier and the KStar classifier (See Figure 5.164). For the Naïve Bayesian classifier, the higher F-measures are obtained with the inclusion of information from each feature's nearest, two-nearest and five-nearest neighbors, then the F-measure drops. The inclusion of neighbor information decreases the F-measures obtained by the MultiBoostAB algorithm using the J48 classifier, with the lower F-measures obtained when information from each feature's nearest and two-nearest neighbors is included. The inclusion of information beyond each feature's nearest neighbors has little effect on the F-measures obtained by the KStar classifier. The F-measures obtained by the KStar are the lowest.


Figure 5.164. F-measure for chromosome fourteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

Like other chromosomes, the inclusion of neighbor information increases the F-measures obtained by the Naïve Bayesian classifier and the KStar classifier for chromosome fifteen. This increase is not observed until the inclusion of information from each feature's five-or-morenearest neighbors for the MultiBoostAB algorithm using the J48 classifier (See Figure 5.165).


Figure 5.165. F-measure for chromosome fifteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome sixteen, the inclusion of neighbor information increases the F-measures obtained by the Naïve Bayesian classifier and the KStar classifier. The inclusion of information from each feature's nearest and two-nearest neighbors decreases the F-measures obtained by the MultiBoostAB algorithm using the J48 classifier. With the inclusion of information from the five- or ten-nearest neighbors of each feature, the F-measures then increase for the MultiBoostAB algorithm using the J48 classifier (See Figure 5.166).


Figure 5.166. F-measure for chromosome sixteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

The F-measures obtained by the MultiBoostAB algorithm using the J48 classifier and the Naïve Bayesian classifier are excellent for all instances for the mitochondrial chromosome. The Fmeasures are low for the KStar classifier, and decrease with the inclusion of neighbor information (See 5.167).


Figure 5.167. F-measure for the mitochondrial chromosome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

The MultiBoostAB algorithm using the J48 classifier and the Naïve Bayesian classifier obtained the higher F-measures for most of the chromosomes. The partitioning of the classification process by chromosome, in general, did not improve the accuracy of the classification for the classifiers.

### 5.2.2. Using Neighbor Gene Ontology Information

A series of experiments was performed to determine if given the properties of chromosome number, start position, stop position, strand, neighbor number, neighbor's strand, neighbor's distance (determined from the mid-point of each feature), neighbor's Gene Ontology aspect, neighbor's Gene Ontology term, Gene Ontology aspect and Gene Ontology term for genomic features (features being: genes, ARS, etc) could classification algorithms correctly classify Gene Ontology term. This was also performed including information from the nearest, two-, five- and ten-nearest neighbors of each feature. This was also repeated on the individual chromosomes including up to ten neighbors.

The results of the experiments are based on averages of ten different runs of each dataset being randomized and then split for training and testing. The training sets were set to contain $66 \%$ of the instances randomly selected from each dataset.

### 5.2.2.1 Results for Entire Genome

When using only a genomic feature's chromosome number, start position, stop position, strand, and Gene Ontology information without neighbor information, we see that the percentage of genomic features correctly classified to the correct Gene Ontology term varies depending on the classification algorithm used (See Figure 5.168). When no neighbor information is included less than $30 \%$ of the genomic features are classified to the correct network regardless of the classifier used. As more neighbor information is added, the KStar classifier and the MultiBoostAB algorithm using the J48 classifier increase their correctly classified percentage with the MultiBoostAB algorithm using the J48 classifier achieving the higher percentages. For the MultiBoostAB algorithm using the J48 classifier, the inclusion of neighbor information beyond each feature's two-nearest neighbors does not impact the percentage of correctly classified features. When the Naïve Bayesian classifier is used, the percentage of correctly classified genomic features does not change with the addition of neighbor information.


Figure 5.168. Percent of genomic features classified for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

When no neighbor information is included, all of the classifiers have poor agreement between the predicted and observed classification (See Figure 5.169). The MultiBoostAB algorithm using the J48 classifier is able to achieve a very good agreement between predicted and observed classification upon the addition of information from each feature's two-nearest neighbors. The Kappa statistic is not altered for the Naïve Bayesian classifier with the inclusion of neighbor information. The inclusion of neighbor information does improve the agreement between the predicted and observed classification for the KStar classifier, but the agreements remain poor.


Figure 5.169. Kappa statistic for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

The inclusion of neighbor information affects the precision of the classification process. For all classifiers, the inclusion of neighbor information improves the classification precision. For the MultiBoostAB algorithm using the J48 classifier, the inclusion of neighbor information beyond each feature's two-nearest neighbors has little effect on the precision. For the KStar classifier, the precision increases as more neighbor information is included, but the precision scores remain low. For the Naïve Bayesian classifier, the inclusion of neighbor information beyond each feature's nearest neighbors has little effect on the precision with the scores leveling out a little above 0.2 (See Figure 5.170).


Figure 5.170. Precision for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For all classifiers, the recall scores are improved by the inclusion of neighbor information. The recall scores obtained by the MultiBoostAB algorithm using the J 48 classifier are the highest, and they increase with the inclusion of more neighbor information. The recall scores of the KStar classifier increases as more neighbor information is included but they remain low. For the Naïve Bayesian classifier, the inclusion of neighbor information beyond each feature's nearest neighbors has little effect on the precision with the scores leveling out around 0.1 (See Figure 5.171).


Figure 5.171. Recall scores for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

F-measures obtained by the different classifiers mimic their precision and recall scores (See Figure 5.172). The accuracy of the classification of a genomic feature increases with the inclusion of neighbor information for all classifiers. The MultiBoostAB algorithm using the J48 classifier obtains the higher F-measures, and the inclusion of neighbor information beyond each feature's two-nearest neighbors has little effect on the F-measures obtained. The accuracy of the KStar classifier increases as more neighbor information is included, but the values remain low. For the Naïve Bayesian classifier, the inclusion of neighbor information beyond each feature's nearest neighbors has little effect on the F-measures obtained.


Figure 5.172. F-measure for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

When analyzing the area under the ROC curves, when no neighbor information is included, the Naïve Bayesian classifier is the only classifier that obtains an accuracy score that is deemed good. This result does not change with the inclusion of neighbor information. For the MultiBoostAB algorithm using the J48 classifier, the accuracy improves from poor to good with the inclusion of each feature's nearest neighbor. Then, it improves to excellent upon the inclusion of information from each feature's two-nearest neighbors. The KStar classifier improves in a similar manner, but its scores are not as high as those that are obtained using the MultiBoostAB algorithm using the J48 classifier (See Figure 5.173).


Figure 5.173. Area under ROC curve for the entire genome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

Based on these results, the MultiBoostAB algorithm using the J 48 classifier performs the best, achieving significantly higher scores in all of the metrics used.

To see the effect on the accuracy of classification for each Gene Ontology term, the F-measures obtained by the MultiBoostAB algorithm using the J48 classifier for each Gene Ontology term were analyzed. The table of the F-measures can be found in Appendix F. Most of the Gene Ontology term groups benefited from the addition of neighbor information (See Figures 5.174 and 5.175). Based on the results, the inclusion of neighbor information beyond each feature's two-nearest neighbors did not significantly increase the accuracy of classification of features to their correct Gene Ontology terms. When no neighbor information is included, Gene Ontology terms that represent broader functionalities (e.g. biological process, molecular function) have higher accuracy scores, meaning the classifier was able to classify features that belonged to these Gene Ontology terms better. The four Gene Ontology term groups that obtained the lowest F-
measures (did not achieve an F-measure above 0.5) are: anatomical structure morphogenesis, cell cortex, cellular bud and site of polarized growth.


Figure 5.174. F-measures obtained using the MultiBoostAB algorithm using the J48 classifier for the individual Gene Ontology (GO) terms.


Figure 5.175. F-measures obtained using the MultiBoostAB algorithm using the J48 classifier for the individual Gene Ontology (GO) terms.

### 5.2.2.2 Results for Individual Chromosomes

When we look at each chromosome individually, we see that the results of classification vary from chromosome to chromosome, from the number of neighbors included, and from the classification algorithm being used.

### 5.2.2.2.1 Percent Correctly Classified

For all chromosomes, the percent of genomic features correctly classified to the correct Gene Ontology term differ depending on the classifier used and the amount of neighbor information included.

For chromosome one, the inclusion of neighbor information increases the percentage of correctly classified features for the classifiers (See 5.176). The larger increases are seen in the MultiBoostAB algorithm using the J48 classifier, with this classifier obtaining over $80 \%$ correctly classified features after the inclusion of information from each feature's two-nearest neighbors. The KStar classifier obtains the lowest percentages of correctly classified features.


Figure 5.176. Percent of genomic features classified for chromosome one using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome two, the MultiBoostAB algorithm using the J48 classifier and the KStar classifier obtain increases in the percentage of correctly classified features with the inclusion of neighbor information with the MultiBoost AB algorithm using the J48 classifier achieving the higher percentages. The Naïve Bayesian classifier is not affected by the inclusion of neighbor information (See Figure 5.177).

5.177. Percent of genomic features classified for chromosome two using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome three, the inclusion of neighbor information increases the percentage of correctly classified features for all of the classifiers. The MultiBoostAB algorithm using the J48 classifier obtains over $80 \%$ correctly classified features after the inclusion of information from the two-nearest neighbors of each feature (See Figure 5.178).


Figure 5.178. Percent of genomic features classified for chromosome three using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome four, the MultiBoostAB algorithm using the J48 classifier and the KStar classifiers are the most affected by the inclusion of neighbor information. The MultiBoostAB algorithm using the J48 classifier obtains the higher percentages after the inclusion of each feature's two-nearest neighbors. The KStar classifier obtains the lowest percentages of correctly classified features. The Naïve Bayesian classifier is not affected by the inclusion of neighbor information (See Figure 5.179).


Figure 5.179. Percent of genomic features classified for chromosome four using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome five, similar to chromosomes two and four the MultiBoostAB algorithm using the J48 classifier and the KStar classifiers are the most affected by the inclusion of neighbor information The MultiBoostAB algorithm using the J48 classifier obtains the higher percentages, and the KStar classifier obtains the lower percentages. The Naïve Bayesian classifier is not affected by the inclusion of neighbor information, but obtains higher percentages of correctly classified features than the KStar classifier (See Figure 5.180).


Figure 5.180. Percent of genomic features classified for chromosome five using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome six, the inclusion of neighbor information increases the percentage of correctly classified features for all of the classifiers. The MultiBoostAB algorithm using the J48 classifier obtains the higher percentage of correctly classified features, followed by the Naïve Bayesian classifier then the KStar classifier (See Figure 5.181).


Figure 5.181. Percent of genomic features classified for chromosome six using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

Similar to other chromosomes, for chromosome seven, the MultiBoostAB algorithm using the J48 classifier and the KStar classifiers are the most affected by the inclusion of neighbor information. The MultiBoostAB algorithm using the J 48 classifier obtains the higher percentages after the inclusion of each feature's two-nearest neighbors. The KStar classifier obtains the lowest percentages of correctly classified features. The Naïve Bayesian classifier is not affected by the inclusion of neighbor information, but obtains higher percentages than the KStar classifier (See Figure 5.182).


Figure 5.182. Percent of genomic features classified for chromosome seven using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome eight, the patterns of the percentage of correctly classified features is similar to those seen in other chromosomes with the MultiBoostAB algorithm using the J48 classifier and the KStar classifiers being the most affected by the inclusion of neighbor information. The MultiBoost AB algorithm using the J 48 classifier obtains the higher percentages after the inclusion of each feature's two-nearest neighbors. The KStar classifier obtains the lowest percentages of correctly classified features (See Figure 5.183).


Figure 5.183. Percent of genomic features classified for chromosome eight using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

Chromosome nine repeats patterns seen in prior chromosomes. The MultiBoostAB algorithm using the J48 classifier obtains the higher percentages after the inclusion of each feature's twonearest neighbors. The KStar classifier obtains the lower percentages of correctly classified features. The Naïve Bayesian classifier is not affected by the inclusion of neighbor information (See Figure 5.184).


Figure 5.184. Percent of genomic features classified for chromosome nine using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

Similar to other chromosomes, for chromosome ten the MultiBoostAB algorithm using the J48 classifier obtains the higher percentages after the inclusion of each feature's two-nearest neighbors. The KStar classifier obtains the lower percentages of correctly classified features. The Naïve Bayesian classifier is not affected by the inclusion of neighbor information (See Figure 5.185).


Figure 5.185. Percent of genomic features classified for chromosome ten using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome eleven, the inclusion of neighbor information increases the percentage of correctly classified features for the MultiBoostAB algorithm using the J48 classifier and the KStar classifier (See Figure 5.186). The Naïve Bayesian classifier is not affected by the inclusion of neighbor information. The MultiBoostAB algorithm using the J48 classifier obtains the higher percentages of correctly classified features, and the inclusion of neighbor information beyond each feature's two-nearest neighbors has little impact on the percentage of correctly classified features.


Figure 5.186. Percent of genomic features classified for chromosome eleven using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome twelve, the inclusion of neighbor information increases the percentage of correctly classified features for the MultiBoostAB algorithm using the J48 classifier and the KStar classifier. The Naïve Bayesian classifier is not affected by the inclusion of neighbor information (See Figure 5.187). The MultiBoostAB algorithm using the J48 classifier obtains the higher percentages of correctly classified features, and the inclusion of neighbor information beyond each feature's two-nearest neighbors has little impact on the percentage of correctly classified features.


Figure 5.187. Percent of genomic features classified for chromosome twelve using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

Similar to other chromosomes, for chromosome thirteen, the inclusion of neighbor information increases the percentage of correctly classified features for the MultiBoostAB algorithm using the J48 classifier and the KStar classifier. The Naïve Bayesian classifier is not affected by the inclusion of neighbor information (See 5.188). The MultiBoostAB algorithm using the J48 classifier obtains the higher percentages of correctly classified features, and the inclusion of neighbor information beyond each feature's two-nearest neighbors has little impact on the percentage of correctly classified features.


Figure 5.188. Percent of genomic features classified for chromosome thirteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome fourteen, the inclusion of neighbor information increases the percentage of correctly classified features for the MultiBoostAB algorithm using the J48 classifier and the KStar classifier. For the Naïve Bayesian classifier, the percentage of correctly classified features decreases with the inclusion of neighbor information (See 5.189). The MultiBoostAB algorithm using the J48 classifier obtains the higher percentages of correctly classified features, and the inclusion of neighbor information beyond each feature's two-nearest neighbors has little impact on the percentage of correctly classified features. The inclusion of neighbor information beyond each feature's nearest neighbors has little impact on the percentage of correctly classified features for the KStar classifiers.


Figure 5.189. Percent of genomic features classified for chromosome fourteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome fifteen, like other chromosomes, the inclusion of neighbor information increases the percentage of correctly classified features for the MultiBoostAB algorithm using the J48 classifier and the KStar classifier. The Naïve Bayesian classifier is not affected by the inclusion of neighbor information (See 5.190). The MultiBoostAB algorithm using the J48 classifier obtains the higher percentages of correctly classified features, and the inclusion of neighbor information beyond each feature's two-nearest neighbors has little impact on the percentage of correctly classified features.


Figure 5.190. Percent of genomic features classified for chromosome fifteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome sixteen, similar to other chromosomes, the inclusion of neighbor information increases the percentage of correctly classified features for the MultiBoostAB algorithm using the J48 classifier and the KStar classifier. The Naïve Bayesian classifier is not affected by the inclusion of neighbor information (See 5.191). The MultiBoostAB algorithm using the J48 classifier obtains the higher percentages of correctly classified features, and the inclusion of neighbor information beyond each feature's two-nearest neighbors has little impact on the percentage of correctly classified features. The inclusion of neighbor information from each feature's five-nearest neighbors obtains the highest F-measure for the KStar classifier.


Figure 5.191. Percent of genomic features classified for chromosome sixteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For the mitochondrial chromosome, the inclusion of neighbor information increases the percentage of correctly classified features for the MultiBoostAB algorithm using the J48 classifier and the Naïve Bayesian classifier (See 5.192). The MultiBoostAB algorithm using the J48 classifier obtains the higher percentages of correctly classified features, and the inclusion of neighbor information beyond each feature's nearest neighbors has little impact on the percentage of correctly classified features. The inclusion of neighbor information from each feature's fivenearest neighbors obtains the highest F-measure for the Naïve Bayesian classifier. For the KStar classifier, the inclusion of neighbor information from each feature's nearest, two-nearest, or tennearest neighbors does increase the percentage of correctly classified features.


Figure 5.192. Percent of genomic features classified for the mitochondrial chromosome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

The MultiBoostAB algorithm using the J48 classifier obtains the higher percentages of correctly classified features for all of the chromosomes. For most chromosomes, the inclusion of neighbor information beyond each feature's two-nearest neighbors did not have a huge effect on the percentage of correctly classified features. The partitioning of the classification process by chromosome slightly improves the percentage of correctly classified features for MultiBoostAB algorithm using the J48 classifier.

### 5.2.2.2.2 F-measures

To look at the accuracy of the classification process for each chromosome, the F-measures obtained for each chromosome can be analyzed.

For chromosome one, the inclusion of neighbor information is beneficial, increasing the Fmeasures obtained by all of the classifiers (See Figure 5.193). The F-measures are higher for the MultiBoostAB algorithm using the J 48 classifier and lower for the KStar classifier. For the

MultiBoostAB algorithm using the J48 classifier, including neighbor information beyond the two-nearest neighbors of each feature does not add an increased benefit to the classification process.


Figure 5.193. F-measure for chromosome one using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome two, the inclusion of neighbor information increases the F-measures obtained by the KStar classifier and the MultiBoostAB algorithm using the J48 classifier. Including neighbor information has little effect on the Naïve Bayesian classifier (See Figure 5.194). The higher F-measures are obtained by the MultiBoostAB algorithm using the J48 classifier. The addition of neighbor information beyond each feature's two-nearest neighbors does not add an increased benefit to the classification process for the MultiBoostAB algorithm using the J48 classifier.


Figure 5.194. F-measure for chromosome two using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome three, the inclusion of neighbor information is beneficial, increasing the Fmeasures obtained by all of the classifiers (See Figure 5.195). The MultiBoostAB algorithm using the J48 classifier obtains the higher F-measures, with the inclusion of information from each feature's five-nearest neighbors obtaining the highest F-measure.


Figure 5.195. F-measure for chromosome three using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome four, the inclusion of neighbor information increases the F-measures obtained by the KStar classifier and the MultiBoostAB algorithm using the J48 classifier. Including neighbor information has little effect on the Naïve Bayesian classifier (See Figure 5.196). The higher F-measures are obtained by the MultiBoostAB algorithm using the J 48 classifier. The addition of neighbor information beyond each feature's five-nearest neighbors does not add an increased benefit to the classification process for the MultiBoostAB algorithm using the J48 classifier.


Figure 5.196. F-measure for chromosome four using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome five, the inclusion of neighbor information is beneficial, increasing the Fmeasures obtained by all of the classifiers (See Figure 5.197). The MultiBoostAB algorithm using the J48 classifier obtains the higher F-measures. The addition of neighbor information beyond each feature's two-nearest neighbors does not add an increased benefit to the classification
process for the MultiBoostAB algorithm using the J48 classifier. The Naïve Bayesian classifier obtains the lower F-measures.


Figure 5.197. F-measure for chromosome five using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome six, the inclusion of neighbor information is beneficial, increasing the Fmeasures obtained by all of the classifiers (See Figure 5.198). The MultiBoostAB algorithm using the J48 classifier obtains the higher F-measures. The inclusion of neighbor information beyond each feature's two-nearest neighbors does not add an increased benefit to the classification process for the MultiBoostAB algorithm using the J48 classifier. The KStar classifier obtains the lower F-measures until the addition of information from each feature's ten-nearest neighbors when it surpasses the results obtained by the Naïve Bayesian classifier.


Figure 5.198. F-measure for chromosome six using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome seven, the inclusion of neighbor information is beneficial, increasing the Fmeasures obtained by all of the classifiers (See Figure 5.199). The MultiBoostAB algorithm using the J48 classifier obtains the higher F-measures. The inclusion of neighbor information beyond each feature's two-nearest neighbors does not add an increased benefit to the classification process for the MultiBoostAB algorithm using the J48 classifier. For the Naïve Bayesian classifier, the inclusion of neighbor information beyond each feature's nearest neighbors does not add an increased benefit to the classification process. The KStar classifier obtains the lower F-measures until the addition of information from each feature's two-nearest neighbors when it surpasses the results obtained by the Naïve Bayesian classifier.


Figure 5.199. F-measure for chromosome seven using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome eight, the inclusion of neighbor information is beneficial, increasing the Fmeasures obtained by the KStar classifier and the MultiBoostAB algorithm using the J48 classifier. Including neighbor information has little effect on the Naïve Bayesian classifier (See Figure 5.200). The MultiBoostAB algorithm using the J48 classifier obtains the higher Fmeasures. The inclusion of neighbor information beyond each feature's two-nearest neighbors does not add an increased benefit to the classification process for the MultiBoostAB algorithm using the J48 classifier. The KStar classifier obtains the lower F-measures until the addition of information from each feature's two-nearest neighbors when it surpasses the results obtained by the Naïve Bayesian classifier.


Figure 5.200. F-measure for chromosome eight using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome nine, the inclusion of neighbor information is beneficial, increasing the Fmeasures obtained by all of the classifiers (See Figure 5.201). The MultiBoostAB algorithm using the J48 classifier obtains the higher F-measures. The inclusion of neighbor information beyond each feature's two-nearest neighbors does not add an increased benefit to the classification process for the MultiBoostAB algorithm using the J48 classifier. The Naïve Bayesian classifier tends to obtain the lower F-measures.


Figure 5.201. F-measure for chromosome nine using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome ten, the inclusion of neighbor information is beneficial, increasing the Fmeasures obtained by all of the classifiers (See Figure 5.202). The MultiBoostAB algorithm using the J48 classifier obtains the higher F-measures. The inclusion of neighbor information beyond each feature's two-nearest neighbors does not add an increased benefit to the classification process for the MultiBoostAB algorithm using the J48 classifier. For the Naïve Bayesian classifier, the inclusion of neighbor information beyond each feature's nearest neighbors does not add an increased benefit to the classification process. The KStar classifier obtains the lower F-measures until the addition of information from each feature's two-nearest neighbors when it surpasses the results obtained by the Naïve Bayesian classifier.


Figure 5.202. F-measure for chromosome ten using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome eleven, the inclusion of neighbor information is beneficial, increasing the Fmeasures obtained by all of the classifiers, but the increases are small for the Naïve Bayes classifier (See Figure 5.203). The MultiBoostAB algorithm using the J48 classifier obtains the higher F-measures. The inclusion of neighbor information beyond each feature's two-nearest neighbors does not add an increased benefit to the classification process for the MultiBoostAB algorithm using the J48 classifier. In general, the Naïve Bayesian classifier obtains the lower Fmeasures.


Figure 5.203. F-measure for chromosome eleven using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome twelve, the inclusion of neighbor information is beneficial, increasing the Fmeasures obtained by all of the classifiers (See Figure 5.204). The MultiBoostAB algorithm using the J48 classifier obtains the higher F-measures. The inclusion of neighbor information beyond each feature's two-nearest neighbors does not add an increased benefit to the classification process for the MultiBoostAB algorithm using the J48 classifier. For the Naïve Bayesian classifier, the inclusion of neighbor information beyond each feature's nearest neighbors does not add an increased benefit to the classification process. The KStar classifier obtains the lower F-measures until the addition of information from each feature's two-nearest neighbors when it surpasses the results obtained by the Naïve Bayesian classifier.


Figure 5.204. F-measure for chromosome twelve using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome thirteen, the inclusion of neighbor information is beneficial, increasing the Fmeasures obtained by the KStar classifier and the MultiBoostAB algorithm using the J48 classifier. Including neighbor information has little effect on the Naïve Bayesian classifier (See Figure 5.205). The MultiBoostAB algorithm using the J48 classifier obtains the higher Fmeasures. The inclusion of neighbor information beyond each feature's two-nearest neighbors does not add an increased benefit to the classification process for the MultiBoostAB algorithm using the J48 classifier. The Naïve Bayesian classifier tends to get the lower F-measures.


Figure 5.205. F-measure for chromosome thirteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome fourteen, the inclusion of neighbor information is beneficial, increasing the Fmeasures obtained by the KStar classifier and the MultiBoostAB algorithm using the J48 classifier (See Figure 5.206). The MultiBoostAB algorithm using the J48 classifier obtains the higher F-measures. The inclusion of neighbor information beyond each feature's two-nearest neighbors does not add an increased benefit to the classification process for the MultiBoostAB algorithm using the J48 classifier. The inclusion of neighbor information beyond each feature's nearest neighbors does not add an increased benefit to the classification process for the KStar classifier. The inclusion of information from each feature's five-nearest neighbors has the most impact on the Naïve Bayesian classifier.


Figure 5.206. F-measure for chromosome fourteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome fifteen, the inclusion of neighbor information is beneficial, increasing the Fmeasures obtained by all of the classifiers (See Figure 5.207). The MultiBoostAB algorithm using the J48 classifier obtains the higher F-measures. The inclusion of neighbor information beyond each feature's two-nearest neighbors does not add an increased benefit to the classification process for the MultiBoostAB algorithm using the J48 classifier. For the Naïve Bayesian classifier, the inclusion of neighbor information beyond each feature's nearest neighbors does not add an increased benefit to the classification process, and it obtains the lower F-measures.


Figure 5.207. F-measure for chromosome fifteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For chromosome sixteen, the inclusion of neighbor information is beneficial, increasing the Fmeasures obtained by all of the classifiers (See Figure 5.208). The MultiBoostAB algorithm using the J48 classifier obtains the higher F-measures. The inclusion of neighbor information beyond each feature's two-nearest neighbors does not add an increased benefit to the classification process for the MultiBoostAB algorithm using the J48 classifier. In fact, including information from each feature's ten-nearest neighbors sees a drop in F-measures. For the Naïve Bayesian classifier, the inclusion of neighbor information beyond each feature's nearest neighbors does not add an increased benefit to the classification process, and it obtains the lower F-measures. For the KStar classifier, the F-measures increase with the addition of neighbor information until the addition of information from each feature's ten-nearest neighbors.


Figure 5.208. F-measure for chromosome sixteen using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

For the mitochondrial chromosome, the MultiBoostAB algorithm using the J48 classifier obtains F-measures of 1.0 independent of the inclusion of neighbor information. The Naïve Bayesian classifier performs in a similar manner to the MultiBoostAB algorithm using the J 48 classifier. For the KStar classifier, the inclusion of neighbor information from each feature's nearest, twonearest, or ten-nearest neighbors is beneficial to the classification process (See Figure 5.209).


Figure 5.209. F-measure for the mitochondrial chromosome using KStar, MultiBoostAB using J48 and Naïve Bayes algorithms for classification.

The F-measures obtained by the MultiBoostAB algorithm using the J48 classifier are higher for all of the chromosomes indicating a higher accuracy of its classification process compared to the other classifiers. For most chromosomes, the inclusion of neighbor information beyond each feature's two-nearest neighbors did not add an increased benefit to the classification process for the MultiBoostAB algorithm using the J48 classifier. The partitioning of the classification process by chromosome, in general, produced a higher accuracy of classification for the MultiBoostAB algorithm using the J48 classifier. The decision trees generated can be found in Appendix I.

### 5.2.2.3 Removing Location Information

A series of experiments was performed to determine if given the properties of neighbor number, neighbor's Gene Ontology information, neighbor's strand, neighbor's distance (determined from the mid-point of each feature), and Gene Ontology information for genomic features (features being: genes, $A R S$, etc) for training could the MultiBoostAB algorithm using the J48 classifier
could correctly classify Gene Ontology term. The MultiBoostAB algorithm using the J48 classifier was selected since it obtained the best performance in the prior section.

These experiments were performed to see if rules could be generated that would provided insight about a genomic feature's network involvement based only on what other features are near it regardless of that feature's location.

The rules generated did not incorporate the neighbor's Gene Ontology information. Neighbor distance was the most important attribute being selected for the root of the decision tree (See Figure 5.210). Increasing the amount of information did not alter the performance of the classification process.


Figure 5.210. The percentage of correctly classified features using the MultiBoostAB algorithm using the J48 classifier.

The accuracy of the classification was low regardless of the number of neighbors included, but there is a slight increase in the F-measure with the inclusion of more neighbor information (See Figure 5.211).


Figure 5.211. The F-measure using the MultiBoostAB algorithm using the J 48 classifier.

### 5.3 Suggesting Locations

To see if locations of interest could be suggested given feature descriptions, different sets of experiments were performed.

### 5.3.1 Classifying Genomic Features to Chromosome

A series of experiments was performed to determine whether genomic features can be classified to the correct chromosome if given different properties. When given the training data of a genomic feature's chromosome number and either its network involvement from KEGG or its Gene Ontology term, the classification results are similar.

The results of the experiments are based on averages of ten different runs of each dataset being randomized and then split for training and testing. The training sets were set to contain $66 \%$ of the instances randomly selected from each dataset.

### 5.3.1.1 Percent of Correctly Classified

The percent of correctly classified features does not vary depending on the classifier used and the information included (See Figure 5.212). The addition of either the Gene Ontology information or network involvement does not make a significant impact on the classification process.


Figure 5.212. The overall correctly classified features for the KStar classifier, the MultiBoostAB algorithm using the J48 classifier and the Naïve Bayes classifier with either KEGG network involvement included or the GO term included.

### 5.3.1.2 Effect on Individual Chromosomes

For a majority of chromosomes, the higher F-measures are obtained using the Naïve Bayesian classifier (See Figure 5.213). The classifiers tend to classify all features to chromosome four (the largest chromosome), and these classifications then show not very high accuracy yielding Fmeasures less than 0.25 . The mitochondrial chromosome only has features that are involved in one network, so that explains the F-measure observed when using information from each feature's network involvement.


Figure 5.213. The F-measures for the KStar classifier, the MultiBoostAB algorithm using the J48 classifier and the Naïve Bayes classifier with either KEGG network involvement included or the GO term included for each chromosome.

### 5.3.2 Classifying Genomic Features to Start Position Using Network Involvement

Experiments were performed to see if given the chromosome number, network involvement and discretized start position to train on that the classifiers could correctly classify the features to the correct start position.

All of the classifiers performed in a similar manner. This can be seen by examining the percentage of correctly classified features (See Table 5.1)

|  | KStar | MultiBoostAB using <br> J48 | Naïve Bayes |
| :--- | :--- | :--- | :--- |
| Percent Correctly <br> Classified | $22.91 \%$ | $22.61 \%$ | $22.59 \%$ |

Table 5.1. Percentage of correctly classified features to start position given chromosome number, network involvement and start position to train on using the KStar classifier, the MultiBoostAB algorithm using the J48 classifier and the Naïve Bayesian classifier on the entire genome.

The rules generated from these experiments tended to clump all the features to one area on the chromosome. Network involvement was not an attribute that was included in any of the rules.

Table 5. 2 represents the rules generated by the MultiBoostAB algorithm using the J48 classifier.
The percentage of correctly classified features to the range parallels the percentage of the chromosome the range represents.

| If on Chromosome | Then start position <br> range | Percent of correctly <br> classified features by <br> rules | Percent of <br> Chromosome <br> covered by range |
| :--- | :--- | ---: | ---: |
| One | $(0-153220.1]$ | 68.05 | 66.56 |
| Two | $(0-153220.1]$ | 20.11 | 18.84 |
| Three | $(0-153220.1]$ | 52.88 | 48.39 |
| Four | $(306406.2-459592.3]$ | 11.34 | 10.00 |
| Five | $(0-153220.1]$ | 27.42 | 26.56 |
| Six | $(0-153220.1]$ | 54.77 | 56.72 |
| Seven | $(306406.2-459592.3]$ | 15.99 | 14.04 |
| Eight | $(0-153220.1]$ | 29.77 | 27.23 |
| Nine | $(153220.1-306406.2]$ | 37.10 | 34.82 |
| Ten | $(0-153220.1]$ | 22.61 | 20.55 |
| Eleven | $(153220.1-306406.2]$ | 27.71 | 22.99 |
| Twelve | $(459592.3-612778.4]$ | 18.50 | 14.21 |
| Thirteen | $(153220.1-306406.2]$ | 18.57 | 16.57 |
| Fourteen | $(306406.2-459592.3]$ | 21.92 | 19.53 |
| Fifteen | $(765964.5-919150.6]$ | 15.45 | 14.04 |
| Sixteen | $(765964.5-919150.6]$ | 18.17 | 16.16 |
| Mitochondrial | $(0-153220.1]$ | 100.00 | 100.00 |

Table 5.2. A representation of the rules generated by the MultiBoostAB algorithm using the J48 classifier. The percent of correctly classified features for each rule was determined by dividing the number of features that were correctly classified by the rule by the total number of features that were classified by that rule then multiplying by 100 .

A series of experiments was performed to determine if partitioning the classification to the individual chromosomes would impact the classification process. The classifiers were trained on each feature's network involvement and discretized start position. Each of the classifiers performed similarly (See Table 5.3).

| Chromosome | KStar | MultiBoostAB using <br> J48 | Naïve Bayes |
| :--- | :--- | :--- | :--- |
| One | $13.74 \%$ | $13.74 \%$ | $14.70 \%$ |
| Two | $11.71 \%$ | $11.19 \%$ | $12.15 \%$ |
| Three | $14.66 \%$ | $13.54 \%$ | $15.40 \%$ |
| Four | $11.49 \%$ | $11.34 \%$ | $10.91 \%$ |
| Five | $14.35 \%$ | $13.18 \%$ | $13.19 \%$ |
| Six | $12.71 \%$ | $12.22 \%$ | $12.47 \%$ |
| Seven | $11.91 \%$ | $11.45 \%$ | $11.84 \%$ |
| Eight | $14.09 \%$ | $13.37 \%$ | $14.09 \%$ |
| Nine | $12.74 \%$ | $12.58 \%$ | $12.90 \%$ |
| Ten | $12.98 \%$ | $12.60 \%$ | $12.88 \%$ |
| Eleven | $11.78 \%$ | $12.36 \%$ | $11.43 \%$ |
| Twelve | $14.82 \%$ | $14.82 \%$ | $15.80 \%$ |
| Thirteen | $10.49 \%$ | $10.96 \%$ | $10.80 \%$ |
| Fourteen | $12.49 \%$ | $11.47 \%$ | $13.04 \%$ |
| Fifteen | $11.72 \%$ | $11.59 \%$ | $11.99 \%$ |
| Sixteen | $13.02 \%$ | $12.72 \%$ | $13.40 \%$ |
| Mitochondrial | $17.24 \%$ | $17.24 \%$ | $17.24 \%$ |

Table 5.3.Percentage of correctly classified features to start position given network involvement and start position to train on using the KStar classifier, the MultiBoostAB algorithm using the J48 classifier and the Naïve Bayesian classifier on the individual chromosomes.

The results from above show that if a feature's network involvement is known, it would be hard to accurately determine which chromosome it is located on. Even if the chromosome number is determined, the classifiers then only have anywhere between a $10-17 \%$ chance of locating the feature's starting position.

The classification process was performed using information from the entire genome (each feature's network involvement and start position but no chromosome number information to train on) to see if features of certain networks could be attributed to certain areas on any chromosome. The classifiers performed in a similar manner (See Table 5.4)

|  | KStar | MultiBoostAB using <br> J48 | Nä̈ve Bayes |
| :--- | :--- | :--- | :--- |
| Percent Correctly <br> Classified | $21.05 \%$ | $21.05 \%$ | $21.09 \%$ |

Table 5.4. Percentage of correctly classified features to start position given network involvement and start information to train on using the KStar classifier, the MultiBoostAB algorithm using the J48 classifier and the Naïve Bayesian classifier for the entire genome.

### 5.3.2.1Generating Rules to Classify Start Position Using Network Involvement

## Information

Since network involvement was not an important attribute in generating rules for the MultiBoostAB algorithm using the J48 classifier, other rules generating classifiers were investigated. The Nearest-neighbor-like algorithm using non-nested generalized exemplars (which are hyperrectangles that can be viewed as if-then rules) obtained rules that did use network involvement to classify start position. The percentage of correctly classified features to their start position for this classifier is $22.653 \%$. The rules that were generated were adapted and used with the IR code to make suggestions of where to look for networks (See section 5.3.4). The rule adaption is discussed in Appendix G. The suggestions of the rules tended to be correct, but they neglected to suggest all the locations that actually contained features that participated in the given network.

### 5.3.3 Classifying Genomic Features to Start Position Using Gene Ontology Term

Experiments were performed to see if given the chromosome number, the feature's Gene Ontology term and discretized start position to train on that the classifiers could correctly classify the features to the correct start position. Table 5.5 contains the percent of correctly classified features for the different classifiers.

|  | KStar | MultiBoostAB using <br> J48 | Naïve Bayes |
| :--- | :--- | :--- | :--- |
| Percent Correctly <br> Classified | $22.24 \%$ | $21.60 \%$ | $22.59 \%$ |

Table 5.5. Percentage of correctly classified features to start position given chromosome number, network involvement and start position to train on using the KStar classifier, the MultiBoostAB algorithm using the J48 classifier and the Naïve Bayesian classifier on the entire genome.

The rules generated from these experiments also tended to clump all the features to one area on the chromosome for some chromosomes.

A series of experiments was performed to determine if partitioning the classification to the individual chromosomes would impact the classification process. The classifiers were trained on each feature's Gene Ontology term and discretized start position for each chromosome separately. Each of the classifiers performed similarly (See Table 5.6). The Gene Ontology term was an attribute in the rules generated for the MultiBoostAB algorithm using the J48 classifier.

| Chromosome | KStar | MultiBoostAB using <br> J48 | Naïve Bayes |
| :--- | :--- | :--- | :--- |
| One | $13.67 \%$ | $12.17 \%$ | $12.17 \%$ |
| Two | $10.49 \%$ | $11.30 \%$ | $11.30 \%$ |
| Three | $11.81 \%$ | $12.87 \%$ | $12.55 \%$ |
| Four | $10.05 \%$ | $10.33 \%$ | $10.33 \%$ |
| Five | $11.40 \%$ | $10.76 \%$ | $11.15 \%$ |
| Six | $13.82 \%$ | $13.11 \%$ | $12.96 \%$ |
| Seven | $11.03 \%$ | $10.43 \%$ | $10.43 \%$ |
| Eight | $12.13 \%$ | $11.60 \%$ | $11.60 \%$ |
| Nine | $11.47 \%$ | $10.33 \%$ | $11.03 \%$ |
| Ten | $11.81 \%$ | $11.71 \%$ | $11.92 \%$ |
| Eleven | $10.85 \%$ | $11.65 \%$ | $11.28 \%$ |
| Twelve | $13.00 \%$ | $12.43 \%$ | $12.43 \%$ |
| Thirteen | $10.68 \%$ | $9.61 \%$ | $9.53 \%$ |
| Fourteen | $10.44 \%$ | $10.53 \%$ | $10.73 \%$ |
| Fifteen | $12.82 \%$ | $11.16 \%$ | $11.19 \%$ |
| Sixteen | $12.35 \%$ | $12.89 \%$ | $11.52 \%$ |
| Mitochondrial | $16.97 \%$ | $16.97 \%$ | $16.61 \%$ |

Table 5.6. Percentage of correctly classified features to start position given GO term and start position to train on using the KStar classifier, the MultiBoostAB algorithm using the J48 classifier and the Naïve Bayesian classifier on the individual chromosomes.

The results from above show that if a feature's Gene Ontology term is known, it would be hard to accurately determine which chromosome it is located on. Even if the chromosome number is determined, the classifiers then only have anywhere between a $10-16 \%$ chance of locating the feature's starting position.

The classification process was performed using information from the entire genome (each feature's Gene Ontology term and start position but no chromosome number information to
train on) to see if features that have certain Gene Ontology terms could be attributed to certain areas on any chromosome. Each classifier had a similar performance (See Table 5.7).

|  | KStar | MultiBoostAB using <br> J48 | Naïve Bayes |
| :--- | :--- | :--- | :--- |
| Percent Correctly <br> Classified | $21.07 \%$ | $20.57 \%$ | $21.18 \%$ |

Table 5.7. Percentage of correctly classified features to start position given GO term and start information to train on using the KStar classifier, the MultiBoostAB algorithm using the J48 classifier and the Naïve Bayesian classifier for the entire genome.

### 5.3.4 Using Genomic Feature Descriptions and Information Retrieval

A series of experiments was performed to determine whether given a description of a feature, would finding the networks that the top ten similar features (similarity based on IR) were involved in be helpful by suggesting looking by the feature's involved in these networks as areas of possible interest.

For the experiments, descriptions from 8287 genomic features obtained from the Saccharomyces Genome Database were used. For the other 8587 genomic features listed in the Saccharomyces Genome Database, descriptions could not be obtained. For each of these descriptions, the top ten most similar features from the corpus of descriptions were retrieved. To see if these top ten features could properly suggest network involvement for the given feature, the network involvement of each of the top ten features compared to the actual network involvement of the given feature. A tally was kept of how many networks were suggested and of how many of those suggestions were correct.

From the retrieval results of the 8287 features, the percentage of correct suggestions was calculated to be $42.86 \%$. To further analyze how the retrieval process worked, each individual run was analyzed. Table 5.8 shows a summary of the runs

| Number of features | Percentage of features | Description of suggestions |
| :---: | :---: | :---: |
| 4669 | 56.34 | There was no known network, and no suggestions were made |
| 2520 | 30.41 | There was no known network involvement, but suggestions were made |
| 90 | 1.09 | Network suggestion were made, but none of them matched the known network involvement |
| 47 | 0.57 | There was known network involvement, but no suggestions were made. |
| 349 | 4.21 | $100 \%$ of the suggestions were accurate |
| 28 | 0.34 | The range of the percent of correct suggestions was $90 \leq x<100$ |
| 54 | 0.65 | The range of the percent of correct suggestions was $80 \leq x<90$ |
| 42 | 0.51 | The range of the percent of correct suggestions was $70 \leq x<80$ |
| 70 | 0.84 | The range of the percent of correct suggestions was $60 \leq x<70$ |
| 123 | 1.48 | The range of the percent of correct suggestions was $50 \leq x<60$ |
| 67 | 0.81 | The range of the percent of correct suggestions was $40 \leq x<50$ |
| 52 | 0.63 | The range of the percent of correct suggestions was $30 \leq x<40$ |
| 78 | 0.94 | The range of the percent of correct suggestions was $20 \leq x<30$ |
| 75 | 0.91 | The range of the percent of correct suggestions was $10 \leq x<20$ |
| 23 | 0.28 | The range of the percent of correct suggestions was $0<x<10$ |

Table 5.8. Shows a summary of the retrieval results and network suggestions for the 8287 genomic features.

A search engine was developed to take as input a description of a feature. Based on the network involvement of the top ten most similar features to the given feature, locations of start positions in the genome for other features that may participate in a network with the given feature are provided. A diagram of the process can be seen in Figure 3.3. The given feature description does affect the accuracy of the results. The vaguer a description is, the less likely the user will get suggestions that would be useful.

### 5.4 Summary

The results of this study showed that in classifying features to their correct network, neighbor information from those features is needed. The inclusion of the Gene Ontology term for each feature's two-nearest neighbors was more beneficial to the classification process than including either the neighbor's SGDID or the neighbor's network involvement. The MultiBoostAB algorithm using the J48 classifier performed better than the KStar classifier and the Naïve Bayesian classifier. When using the Gene Ontology term for each feature's two-nearest
neighbors, it was determined that partitioning of the classification process improved F-measures obtained by the MultiBoostAB algorithm using the J48 classifier, so the rules that were translated for the rule-based system were based on the rules generated using the individual chromosomes. By doing this, the average percent correctly classified became $50.63 \%$ and the average F-measure became 0.65 . The rules for the rule-based system were fairly accurate and could classify features to their correct network over $50 \%$ of the time. The results did show that knowing the location of the feature is key in classifying it to the proper network. Only using neighbor information to classify a feature to a network reduces the percentage of correctly classified features by over 40 percentage points.

The results of this study also showed that in classifying features to their correct Gene Ontology term group, neighbor information from those features is also needed. The inclusion of the Gene Ontology information for each feature's two-nearest neighbors was more beneficial to the classification process than including the neighbor's SGDID. The MultiBoostAB algorithm using the J48 classifier performed better than the KStar classifier and the Naïve Bayesian classifier. When using the Gene Ontology information for each feature's two-nearest neighbors, it was determined that partitioning of the classification process improved F-measures obtained by the MultiBoostAB algorithm using the J48 classifier, so the rules that were translated for the rulebased system were based on the rules generated using the individual chromosomes. By doing this, the average percent correctly classified became $83.1 \%$ and the average F-measure became 0.85 . The rules for the rule-based system were fairly accurate and could classify features to their correct Gene Ontology term over $80 \%$ of the time. The results did show that knowing the location of the feature is key in classifying it to the proper Gene Ontology term. Only using neighbor information to classify a feature to a network reduces the percentage of correctly classified features by over 50 percentage points.

When trying to suggest locations of features based on their network involvement or their Gene Ontology term, the classification process was not as successful. The locations that were recommended did not include all the known features that belonged to either the network or the Gene Ontology group. These rules only were able to suggest some locations of interest, not all. When trying to suggest areas of interest of possible network work involvement using textual feature descriptions, the search engine created in this study recommended the correct network about $43 \%$ of the time.

## 6 Discussion

### 6.1 Discussion of Results

Through my experiments, I determined what neighbor information was the most beneficial in classifying features to the correct network and the correct Gene Ontology term, and how many neighbors to include in this process for the yeast genome. I found that knowing the Gene Ontology terms of a feature's two-nearest neighbors both upstream and downstream on both strands of a feature is important in determining network involvement of that feature and the Gene Ontology term of that feature. Partitioning the classification process by chromosome allowed the MultiBoostAB algorithm [Webb, 2000] using the J48 classifier [Quinlan, 1993] to provide the most accurate results. The decision trees generated in the classification process became the rules in the rule-based systems I produced. The software created acts as a filter, limiting the search area that one needs to look for other genomic features that are involved in a network with a given feature and suggesting possible functionality for a feature if it is unknown.

It was not too surprising to find that the inclusion of information from only each feature's twonearest neighbors had such an impact on classification since Hurst et al. (2004) found that clusters of co-expressed genes in yeast rarely exceed ten genes or a few kilobases. The inclusion of more neighbors, beyond the two nearest, did not significantly improve the classification process.

Lee and Sonnhammer (2003) found $98 \%$ of metabolic pathways in yeast exhibited significantly higher clustering levels than expected by chance. This suggested that the inclusion of neighbor information would be beneficial in classifying features to networks. When examining Figures 5.48-5.50, we do see that for most networks, classifying features to them accurately is benefitted
from the inclusion of neighbor information supporting the notion set forth by Lee and Sonnhammer (2003). It was hoped that the rule-based system would have been able to classify features to their correct network more than it did (over $50 \%$ of the time) based on the high amount of clustering found in the yeast genome.

The rule-based system designed to suggest a feature's Gene Ontology term performed well. The rules for the rule-based system could classify features to their correct Gene Ontology term over $80 \%$ of the time. Having the inclusion of neighbor's Gene Ontology information being so beneficial in classifying features to the correct Gene Ontology term group supports finding from other studies. In a study by Fukuoka et al., they investigated Gene Ontology categories of gene pairs that were considered highly correlated in chromosomal distance ranges between 0 and 20 kbp and between 980 and 1000 kbp . The results of the pairwise analysis of Gene Ontology category showed that only highly correlated pairs shared the same category and most of these pairs were not duplicates, meaning the genes did not share a common history; this was determined by BLAST analysis [Fukuoka, et al., 2004].

When trying to suggest locations of features based on their network involvement or their Gene Ontology term group, the classification process was not as successful. This is in part that the rules did not cover all the locations, only some of them. When trying to suggest areas of interest of possible network work involvement using textual feature descriptions, the search engine created in this study recommended the correct network about $43 \%$ of the time. Marcotte et al. (1999) found that proteins that are functionally related participate in a common structural complex, metabolic pathway, biological process or closely related physiological function. Since the feature descriptions used contained information about functionality, it was thought that the results of the search engine technique would have been higher. It could be that the results are better. The search engine did make network suggestions that did not match what is currently
known. It could be that the search engine made network involvement suggestions for features, and in future studies in yeast, they could be found relevant.

My experimental results showed that my methodology is able to correctly locate areas of interest for network involvement over $50 \%$ of the time and over $80 \%$ of the time suggest the correct Gene Ontology term(s) for a given genomic feature in yeast. By analyzing data from a different perspective (i.e. focusing on the location and the contribution of the nearest neighbors of each genomic feature), I was able to generate rules that can be used to focus searches for genetic networks or determine gene product functionality. Since other genomes (Homo sapiens, Caenorbabditis elegans, Drosophila melanogaster, Arabidopsis thaliana) exhibit clustering patterns, this methodology should translate to those other genomes. The optimal neighborhood size for each genome would have to be established through experimentation. The classification rules generated in the genomes where we do not know as much as we do in yeast could elucidate new relationships between genes.

If we make the analogy that a genome is a collection of neighborhoods, and a genomic feature is like a house, then knowing your address and what your two-nearest neighbors on both sides of your house and across the street do tells us a lot about what you do (Gene Ontology terms) and what organizations (networks) you might be involved in. The software created in this study takes advantage of this phenomenon making it easier to come up with reliable hypotheses about the location of potential genetic regulatory networks within a genome and about possible functionality of a gene.

### 6.2 Future studies

This dissertation only examined yeast. It would be beneficial to perform the same experiments on other species. This way we could determine if there are common trends in the classification
process among different species. The mouse genome and the human genome are the next genomes of interest because mice are a common animal model for studying human disease.

If we compare the human genome and the mouse genome, it looks like the human chromosomes were chopped up into many pieces and then put back together to form the mouse genome (or it may have occurred the other way). These rearrangements did not occur all at one time, but they occurred over millions of years. Within each of the blocks there are many small changes at the gene level, but what is very interesting to scientists is how much of the DNA sequence is the same between mouse and humans, although the location of the genes is different. This discovery has led scientists to believe that the relative location of genes to one another may also affect the phenotype of an organism [Mouse Genome Sequencing Consortium, 2002]. Discovering patterns in gene location may offer more insight to this.

Also, understanding why genes are located where they are will be useful to scientists to help discover genes that cause disease. The software I developed utilizes patterns based on gene location and the neighborhood around it. Once these patterns have been established in different organisms, then discovering these relationships between animal models and humans will be easier. Many diseases are caused by flawed networks. When genes malfunction, they can cause diseases such as cancer, glaucoma, ADHD, diabetes, epilepsy, asthma, cystic fibrosis and muscular dystrophy. Research is critical for scientists to cultivate a greater understanding of how genes contribute to disease. This understanding and being able to elucidate genetic networks provides the foundation on which new treatments or cures can be developed.

Since results on individual networks and Gene Ontology term groups were obtained, it would be interesting to see why the classifiers could classify features to certain networks or certain Gene Ontology term groups better than others. A more detailed biological analysis would be required.

In a study by Tsai et al. (2007), their findings suggest that transcription factor sharing is not the sole determinant for the co-regulation of adjacent genes in yeast. These studies suggest that yeast have developed a regulatory system which integrates both transcription-factor-based and non-transcription-factor based mechanisms. They believe that the non-transcription-factor based mechanisms may play a greater role in driving co-expression of neighboring genes [Tsai et al., 2007]. Based on this study, it would be interesting to incorporate transcription factor information in studies similar to the ones performed in this thesis to see if there are patterns based on the location genomic features and their neighbors and transcription factor that interact with them.

## 7 References

Aha, D.W., Kibler, D. and Albert, M.K.(1991) Instance-based learning algorithms. Machine Learning, 6:37-66.

Akutsu, T., Miyano, S., Kuhara, S. (2000) Algorithms for inferring qualitative models of biological networks. Pac Symp Biocomput, 293-304.

Altman, DG. (1991)Practical Statistics for Medical Research. London England: Chapman and Hall. page 404.

Altman, R.B., and Raychaudhuri, S. (2001) Whole-genome analysis: challenges beyond clustering. Curr Opin in Struct Biol., 11:340-347.

Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. (1990). Basic local alignment search tool. J Mol Biol, 215(3): 403-410.

Alon, U., Barkai, N., Notterman, D.A., Gish, K., Ybarra, S., Mack, D., and Levine, A.J. (1999) Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc. Natl. Acad. Sci. USA 96:6745-6750.

Balázsi G, Kay KA, Barabási AL, Oltvai ZN. (2003) Spurious spatial periodicity of co-expression in microarray data due to printing design. Nucleic Acids Res., 31(15):4425-33.

Bauer, E., and Kohavi, R. (1998). An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. Macbine Learning.

Ben-Dor, A., R. Shamir, and Z. Yakhini. (1999) Clustering gene expression patterns. J. Comput. Biol., 6:281-297.

Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A., Sampas, N., Dougherty, E., Wang, E., Marincola, F., Gooden, C., Lueders, J., Glatfelter, A., Pollock, P., Carpten, J., Gillanders, E., Leja, D., Dietrich, K., Beaudry, C., Berens, M., Alberts, D., and Sondak, V. (2000) Molecular classification of cutaneous malignant melanoma by gene expression profiling. Nature, 406:536-540.

Blaschke, C., Andrade, M.A., Ouzounis, C. and Valencia, A. (1999) Automatic extraction of biological information from scientific text: protein-protein interactions. Ismb, 15:60-67.

Blumenthal, T. (1998) Gene clusters and polycistronic transcription in eukaryotes. BioEssays, 20:480-487.

Boutanaev, A.M., Kalmykova, A.I., Shevelyov, Y.Y., Nurminsky, D.I. (2002) Large clusters of co-expressed genes in the Drosophila genome. Nature. 2002 Dec 12;420 (6916):666-9.

Brāzma, A., Vilo, J., Ukkonen, E. and Valtonen, K.(1997) Data mining for regulatory elements in yeast genome. Proceeding of the Fifth International Conference on Intelligent Systems for Molecular Biology. AAAI Press. 65--74.

Caron, H., van Schaik, B., van der Mee, M., Baas, F., Riggins, G., van Sluis, P., Hermus, M.C., van Asperen, R., Boon, K., Voute, P.A., Heisterkamp, S., van Kampen, A., Versteeg, R. (2001)

The human transcriptome map: Clustering of highly expressed genes in chromosomal domains. Science, 291:1289-1292.

Chen, T., He, H.L., Church, G.M. (1999) Modeling gene expression with differential equations. Pac Symp Biocomput, 29-40.

Chen, X., Anantha, G., and Wang, X. (2006) An effective structure learning method for constructing gene networks. Bioinformatics, 22(11): 1367-1374.

Cho, R. J., Campbell, M., Winzeler, E., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T., Gabrielian, A., Landsman, D., Lockhart, D. (1998) A genome-wide transcriptional analysis of the mitotic cell cycle. Molecular Cell 2, 65-73.

Clare, A. and King, R.D. (2003) Data mining the yeast genome in a lazy functional language. Practical Aspects of Declarative Languages : 5th International Symposium, PADL 2003, New Orleans, LA, USA. 19-36.

Cleary, J.G., and Leonard E. Trigg, L.E., (1995) K*: An Instance-based Learner Using an Entropic Distance Measure. 12th International Conference on Machine Learning, 108-114.

Cohen, B. A., Mitra, R. D., Hughes, J. D. \& Church, G. M. (2000) A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. Nature Genet. 26, 183-186.

Cooper, G.F. and Herskovits, E. (1992) A Bayesian method for the induction of probabilistic networks from data. Machine Learning, 9:4, 309-347.

Cristianini, N. and Shawe-Taylor, J. (2000) An Introduction to support vector machines and other kernelbased learning methods. Cambridge university Press, Cambridge, United Kingdom.

Dandekar, T., Snel, B., Huynen, M., Bork, P. (1998) Conservation of gene order: a fingerprint of proteins that physically interact. Trends in Biochemical Science, 23:324-328.

Datta, S. and Datta, S. (2003) Comparisons and validation of statistical clustering techniques for microarray gene expression data. Bioinformatics, 19(4):459-466.
"Determining Gene Function from Sequence Information." Rediscovering Biology: Molecular to Global Perspectives. July 14, 2008
[http://www.learner.org/channel/courses/biology/textbook/genom/genom_6.html](http://www.learner.org/channel/courses/biology/textbook/genom/genom_6.html).
D'haeseleer, P., Liang, S., Somogyi, R. (2000) Genetic network inference: from co-expression clustering to reverse engineering. Bioinformatics, 16:707-726.

Dorsett, D. (1999) Distant liaisons: long-range enhancer-promoter interactions in Drosophila. Curr Opin Genet Dev., 9:505-514.

Dyer, C.R. (2001) CS 540 Lecture Notes: Machine Learning. University of Wisconsin, Madison. March 2, 2006 [http://www.cs.wisc.edu/~dyer/cs540/notes/learning.html](http://www.cs.wisc.edu/~dyer/cs540/notes/learning.html).

Eckardt, N.A. (2001) Everything in its place: Conservation of gene order among distantly related plant species. The Plant Cell, 13:723-725.

Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA, 95:14863-14868.

Enault, F., Suhre, K., Poirot, O., Abergel, C., and Claverie, J. M. (2004) Phydbac2: Improved inference of gene function using interactive phylogenomic profiling and chromosomal location analysis. Nucleic Acids Research, 32, Web Server Issue, W336-W339.

Friedman, N., Linial, M., Nachman, I., Pe'er, D.(2000) Using Bayesian networks to analyze expression data. J Comput Biol, 7:601-620.

Friedman, N., Nachman, I. \& Pe'er , D. (1999), Learning Bayesian network structure from massive datasets: The "sparse candidate" algorithm, in `Proc. Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI '99)', pp. 196-205.

Freund, Y. and Schapire, R. E. (1996). Experiments with a new boosting algorithm. Proceedings of the Thirteenth International Conference on Machine Learning (148-156) Bari, Italy: Morgan Kaufmann.

Friedman, R. and Hughes, A.L. (2001) Gene duplication and the structure of eukaryotic genomes. Genome Res. 11: 373-381.

Fukuoka, Y., Inaoka, H., and Kohane, I.S., (2004) Inter-species differences of co-expression of neighboring genes in eukaryotic genomes. BMC Genomics. [http://www.biomedcentral.com/1471-2164/5/4](http://www.biomedcentral.com/1471-2164/5/4).

Gabrielsson, B.L., Carlsson, B. and Carlsson L.M. (2000) Partial genome scale analysis of gene expression in human adipose tissue using DNA array. Obes Res., 8:374-384.

Galperin, M.Y., Koonin, E.V. (2000) Who's your neighbor? New computational approaches for functional genomics. Nature Biotechnology, 18:609-613.

Gene Ontology Consortium (2000) Gene Ontology: tool for the unification of biology. Nature Genetics, 25:25-29.

Gene Ontology Consortium (2004) The Gene Ontology (GO) database and informatics resource. Nucleic Acids Research, 32:D258-261.

Hartuv, E., Schmitt, A., Lange, J., Meier-Ewert, S., Lehrach, H., and Shamir, R. (1999) An algorithm for clustering cDNAs for gene expression analysis using short oligonucleotide fingerprints, Proceedings of the Third International Symposium on Computational Molecular Biology (RECOMB 99). 188-197. ACM Press, New York.

Hong, E.L., Balakrishnan, R., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., Fisk, D.G., Hirschman, J.E., Livestone, M.S., Nash, R., Park, J., Oughtred, R., Skrzypek, M., Starr, B., Theesfeld, C.L., Andrada, R., Binkley, G., Dong, Q., Lane, C., Hitz, B., Miyasato, S., Schroeder, M., Sethuraman, A., Weng, S., Dolinski, K., Botstein, D. and Cherry, J.M. Saccharomyces Genome Database. March, 2006 <http://www.yeastgenome.org/ >.

Hurst, L., Pál, C., and Lercher, M.J., (2004) The Evolutionary Dynamics of Eukaryotic Gene Order. Nature Reviews Genetics, 5:299-310.

Jensen, L.J., Gupta, R., Staerfeldt, H., and Brunak, S., (2003) Prediction of human protein function according to Gene Ontology categories. Bioinformatics, 19:635-642.

Jenssen, T.K., Laegreid, A., Komorowski, J., and Hovig, E. (2001) A literature network of human genes for high-throughput analysis of gene expression. Nat. Genet., 28:21-28.

Kannan, K., N. Amariglio, G. Rechavi, J. Jakob-Hirsch, I. Kela, N. Kaminski, G. Getz, E. Domany, and D. Givol. (2001) DNA microarrays identification of primary and secondary target genes regulated by p53. Oncogene, 20:2225-2234.

Képès, F. (2003) Periodic epi-organization of the yeast genome revealed by the distribution of promoter sites. J. Mol. Biol. 329, 859-865.

Kozak, M. 1999. Initiation of translation in prokaryotes and eukaryotes. Gene, 234: 187-208.
Kruglyak, S. and Tang, H. (2000) Regulation of adjacent yeast genes. Trends Genet. 16, 109-111.
Lee, J.M., and Sonnhammer, E.L.L. (2003) Genomic Gene Clustering Analysis of Pathways in Eukaryotes. Genome Research, 13:875-882.

Lennon, G.S., and Lehrach, H. (1991) Hybridization analysis of arrayed cDNA libraries. Trends Genet., 7:60-75.

Lercher, M.J., Urrutia, A.O., Hurst, L.D. (2002) Clustering of housekeeping genes provides a unified model of gene order in the human genome. Nat Genet., 31:180-183.

Li, Q., Lee, BTK and Zhang, L. (2005) Genome-scale analysis of positional clustering of mouse testis-specific genes. BMC Genomics, 6:7. [http://www.biomedicalcentral.com/14712164/6/7](http://www.biomedicalcentral.com/14712164/6/7).

Mannila, H., Patrikainen, A., Seppanen, J. K. and Kere, J. (2002) Long-range control of expression in yeast. Bioinformatics 18, 482-483.

Marcotte, E. M., Pellegrini, M., Thompson, M. J., Yeates, T. O. \& Eisenberg, D. (1999). A combined algorithm for genome-wide prediction of protein function. Nature, 402:83-86.

Mayor LR, Fleming KP, Muller A, Balding DJ, Sternberg MJE. Clustering of Protein Domains in the Human Genome. J Mol Biol. 2004;340:991-1004.

Michalski, R.S. (1980). Pattern recognition as rule-guided inductive inference. IEEE Transactions on Pattern Analysis and Macbine Intelligence, 2:349-361.

Michalski, R.S., Mozetic, I. Hong, J., and Lavrac, N. (1986). The multipurpose incremental learning system AQ15 and its testing application to three medical domains. Proceedings of the Fifth National Conference on Artificial Intelligence Philadelphia, PA: Morgan Kaufmann. 1041-1045.

Mlecnik, B., Scheideler, M., Hackl, H., and Hartler, J. (2005) PathwayExplorer: web service for visulalizing high-throughput expression data on biological pathways. Nucleic Acids Research. 33, Web Server Issue, W633-W637.

Mouse Genome Sequencing Consortium (2002) Initial sequencing and comparative analysis of the mouse genome Nature 420:520-562.

Niehrs, C. and Pollet, N., (1999) Synexpression groups in eukaryotes. Nature, 402:483-487.
Oliver, B., Parisi, M., and Clark, D., (2002) Gene expression neighborhoods. Journal of Biology, 1:4.

Pál, C. and Hurst, L. D. (2003) Evidence for co-evolution of gene order and recombination rate. Nature Genet. 33, 392-395.

Pandey, R., Guru, R.K. and Mount, D.W. (2004) Pathway Miner: extracting gene association networks from molecular pathways for predicting the biological significance of gene expression microarray data. Bioinformatics, 20(0):1-3.
J. Platt, J. (1998) Fast Training of Support Vector Machines using Sequential Minimal Optimization, in Advances in Kernel Methods - Support Vector Learning, B. Schölkopf, C. Burges, and A. Smola, (Eds.), MIT Press.

Quackenbush, J. (2001) Computational analysis of microarray data. Nat. Rev. Genet., 2:418-427.
Quinlan, J.R. (1979). Discovering rules by induction from large collections of examples. Expert systems in the micro electronic age. D. Michie (Ed.), Edinburgh: Edinburgh University Press.

Quinlan, J.R. (1986). Induction of decision trees. Machine Learning, 1, 81-106.
Quinlan, J.R. (1987). Simplifying decision trees. International Journal of Man-Machine Studies, 27, 221-234.

Quinlan, J.R. (1988). Decision trees and multi-valued attributes. In J.E. Hayes, D. Michie and J. Richards (Eds.), Machine Intelligence 11. Oxford: Oxford University Press.

Quinlan, J.R., and Rivest, R.L. (1989). Inferring decision trees using the Minimum Description Length principle. Information and Computation, 80, 227-248.

Quinlan, J.R. (1990) Learning Logical Definitions from Relations. Machine Learning, 5, 239-266.
Quinlan, J.R. (1993) C4.5: Programs for Machine Learning. San Mateo, CA: Morgan Kauffman.
Raychaudhuri, S., Chang, J.T., Sutphin, P.D. and Altman, R.B. (2002) Associating genes with gene ontology codes using a maximum entropy analysis of biomedical literature. Genome Res, 12:203-214.
Rindflesch, T.C., Tanabe, L., Weinstein, J.N. and Hunter, L. (2000) EDGAR: extraction of drugs, genes and relations from the biomedical literature. Pac. Symp. Biocomput., 5:517-528.

Rose, K. (1998) Deterministic Annealing for Clustering, Compression, Classification, Regression, and Related Optimization Problems, Proceedings of the IEEE, 80:2210-2239.

Roy, P.J., Stuart, J.M., Lund, J., and Kim, S.K. (2002) Chromosomal clustering of muscleexpressed genes in Caenorbabditis elegans. Nature, 418:975-979.

Shannon, Claude E. \& Warren Weaver (1949): A Mathematical Model of Communication. Urbana, IL: University of Illinois Press

Sharan, R., and R. Shamir. (2000) CLICK: A clustering algorithm with applications to gene expression analysis. Proc. 8th International Conference on Intelligent Systems for Molecular Biology (ISMB '00). AAAI Press, Menlo Park, CA. 307-316.

Spellman, P.T. and Rubin, G.M. (2002) Evidence for large domains of similarly expressed genes in the Drosopbila genome. J. Biol., 1: 5.

Spirtes, P., Glymour, C., Scheines, R., Kauffman, S., Aimale, V., and Wimberly, F. (2000) Constructing Bayesian network models for gene expression networks from microarray data. Proceedings of the Atlantic Symposium on Computational Biology, Genome Information Systems and Technology.

Stapley, B.J. and Benoit, G. (2000) Biobibliometrics: information retrieval and visualization from co-occurrences of gene names in Medline abstracts. Pac. Symp. Biocomput., 5:529-540.

Stewart, B., (2002) Predicting project delivery rates using the Naive-Bayes classifier. Predicting project delivery rates using the Naive-Bayes classifier, 14(3): $161-179$.

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E.S. and Golub, T.R. (1999) Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc. Natl. Acad. Sci. USA, 96:2907-2912.

Tape, TG. "The area under an ROC curve." Interpreting Diagnostic Tests. March 28, 2008 [http://gim.unmc.edu/dxtests/roc3.htm](http://gim.unmc.edu/dxtests/roc3.htm).

Tsai, H.K., Su, C.P., Lu, MY.J., Shih, C.H. and Wang, D. (2007) Co-expression of adjacent genes in yeast cannot be simply attributed to shared regulatory system, BMC Genomics. 8:352.
van der Walt, C.M. and Barnard, E. (2006) Data characteristics that determine classifier performance. Proceedings of the Sixteenth Annual Symposium of the Pattern Recognition Association of South Africa. 160-165. Online available at [http://www.merka.org.za/pubs/CvdWalt.pdf](http://www.merka.org.za/pubs/CvdWalt.pdf).
van Someren, E.P., Wessels, L.F.A., Backer, E., and Reinders, M.J.T., (2002) Genetic Network Modeling, Pharmacogenomics. 3(4):507-25.

Wambutt, R., et al. (2000) Progress in Arabidopsis genome sequencing and functional genomics. Journal of Biotechnology, 31:281-292.

Webb, G.I., (2000). MultiBoosting: A Technique for Combining Boosting and Wagging. Machine Learning, 40(2): 159-196, Boston: Kluwer Academic Publishers.

Witten, I.H. and Frank, E. (2005) Data Mining: Practical machine learning tools and techniques, 2nd Edition, Morgan Kaufmann, San Francisco.

Wren, J.D. and Garner, H.R. (2004) Shared relationship analysis: ranking set cohesion and commonalities within a literature-derived relationship network. Bioinformatics, 20:191-198.

Yanai, I., Mellor, J.C., and DeLisi, C., (2002) Identifying functional links between genes using conserved chromosomal proximity. Trends in Genetics, 18:176-179.

Zhang, H., (2004) The Optimality of Naïve Bayes. Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference, Miami Beach, Florida, USA.

## Appendix A: Stop List Used in Information Retrieval

| a | than | ms |
| :--- | :--- | :--- |
| be | when | over |
| had | also | there |
| it | by | with |
| only | her | are |
| she | most | could |
| was | or | inc |
| about | that | mz |
| because | which | s |
| has | an | they |
| its | can | would |
| of | his | as |
| some | mr | for |
| we | other | into |
| after | the | no |
| been | who | so |
| have | any | this |
| last | co | up |
| on | if | at |
| such | mrs | from |
| were | out | is |
| all | their | not |
| but | will | says |
| he | and | to |
| more | corp |  |
| one | in |  |

## Appendix B: F-measures for the networks using KStar classifier

The Table B1 shows the F-measures for the networks using the KStar classifier trained with the properties of chromosome number, start position, stop position, strand, neighbor's number, neighbor's SGDID, neighbor's strand, neighbor's distance (determined from the mid-point of each feature), and network involvement information from KEGG for genomic features (features being: genes, ARS, etc.).

| F-measure | K* |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Network | No | 1 | 2 | 5 | 10 |
| 1- and 2-methylnaphthalene degradation | 0 | 0 | 0 | 0 | 0 |
| 1,4-dichlorobenzene and dicarboxylate metabolism | 0 | 0 | 0 | 0.125 | 0 |
| 3-chloroacrylic acid degradation | 0 | 0 | 0 | 0 | 0 |
| ABC transporters | 0 | 1 | 1 | 1 | 1 |
| alanine and aspartate metabolism | 0 | 0.074 | 0.118 | 0.116 | 0.126 |
| alkaloid biosynthesis II | 0 | 0 | 0.133 | 0.209 | 0.073 |
| alpha-linolenic acid metabolism | 0 | 0 | 0 | 0 | 0 |
| aminoacyl-tRNA biosynthesis | 0.069 | 0.127 | 0.113 | 0.125 | 0.143 |
| aminophosphonate metabolism | 0 | 0 | 0.182 | 0.037 | 0.049 |
| aminosugars metabolism | 0 | 0.414 | 0.649 | 0.66 | 0.68 |
| arachidonic acid metabolism | 0 | 0.353 | 0.167 | 0.282 | 0.237 |
| arginine and proline metabolism | 0 | 0.318 | 0.169 | 0.302 | 0.303 |
| basal transcription factors | 0.2 | 0.625 | 0.827 | 0.972 | 0.967 |
| benzoate degradation via CoA ligation | 0 | 0.039 | 0.042 | 0.021 | 0.03 |
| benzoate degradation via hydroxylation | 0 | 0 | 0 | 0 | 0 |
| beta-alanine metabolism | 0 | 0.061 | 0.048 | 0.065 | 0.073 |
| bile acid biosynthesis | 0 | 0.054 | 0.045 | 0.08 | 0.092 |
| biosynthesis of steroids | 0 | 0.533 | 0.803 | 0.795 | 0.769 |
| biotin metabolism | 0 | 0.471 | 0.545 | 0.478 | 0.663 |
| butanoate metabolism | 0 | 0.068 | 0 | 0.022 | 0.019 |
| C5-branched dibasic acid metabolism | 0 | 0.143 | 0 | 0 | 0 |
| carbon fixation | 0 | 0.042 | 0.042 | 0.051 | 0.058 |
| cell communication | 0 | 0 | 1 | 0.571 | 0.88 |
| cell cycle | 0.123 | 0.586 | 0.758 | 0.767 | 0.789 |
| citrate cycle | 0 | 0.053 | 0.164 | 0.173 | 0.13 |


| cyanoamino acid metabolism | 0 | 0 | 0.113 | 0.036 | 0.031 |
| :---: | :---: | :---: | :---: | :---: | :---: |
| cysteine metabolism | 0 | 0 | 0.045 | 0.016 | 0.017 |
| DNA polymerase | 0 | 0.533 | 0.364 | 0.491 | 0.515 |
| ether lipid metabolism | 0 | 0 | 0 | 0.058 | 0.156 |
| ethylbenzene degradation | 0 | 0 | 0 | 0 | 0 |
| fatty acid biosynthesis | 0 | 0.333 | 0.519 | 0.536 | 0.466 |
| fatty acid elongation in mitochondria | 0 | 0 | 0 | 0 | 0.077 |
| fatty acid metabolism | 0 | 0.089 | 0.179 | 0.201 | 0.264 |
| fluorobenzoate degradation | 0 | 0 | 0 | 0 | 0 |
| folate biosynthesis | 0 | 0.239 | 0.315 | 0.34 | 0.357 |
| fructose and mannose metabolism | 0 | 0.175 | 0.343 | 0.262 | 0.293 |
| gamma-Hexachlorocyclohexane degradation | 0 | 0.25 | 0.295 | 0.196 | 0.249 |
| geraniol degradation | 0 | 0 | 0 | 0 | 0 |
| glutamate metabolism | 0 | 0.088 | 0.175 | 0.128 | 0.102 |
| glutathione metabolism | 0 | 0 | 0.138 | 0.124 | 0.118 |
| glycan structures biosynthesis 1 | 0 | 0.132 | 0.169 | 0.158 | 0.173 |
| glycan structures biosynthesis 2 | 0 | 0.164 | 0.12 | 0.147 | 0.199 |
| glycan structures degradation | 0 | 1 | 1 | 0.889 | 0.941 |
| glycerolipid metabolism | 0 | 0.25 | 0.348 | 0.268 | 0.333 |
| glycerophospholipid metabolism | 0.125 | 0.204 | 0.349 | 0.293 | 0.374 |
| glycine, serine and threonine metabolism | 0 | 0.156 | 0.175 | 0.212 | 0.236 |
| glycolysis/gluconeogenesis | 0 | 0.112 | 0.271 | 0.287 | 0.313 |
| glycosylphosphatidylinositol-anchor biosynthesis | 0 | 0.211 | 0.203 | 0.29 | 0.336 |
| glyoxylate and dicarboxylate metabolism | 0 | 0.14 | 0.273 | 0.178 | 0.242 |
| high-mannose type N -glycan biosynthesis | 0 | 0.162 | 0.229 | 0.258 | 0.282 |
| histidine metabolism | 0 | 0.211 | 0.337 | 0.3 | 0.327 |
| inositol phosphate metabolism | 0 | 0.075 | 0.091 | 0.109 | 0.136 |
| limonene and pinene degradation | 0 | 0 | 0 | 0 | 0.04 |
| lipoic acid metabolism | 0 | 0.4 | 0.667 | 0.9 | 1 |
| lysine biosynthesis | 0 | 0.245 | 0.318 | 0.277 | 0.376 |
| lysine degradation | 0 | 0.103 | 0.111 | 0.082 | 0.103 |
| MAPK signaling pathway | 0 | 0.622 | 0.691 | 0.735 | 0.748 |
| metabolism of xenobiotics by cytochrome P450 | 0 | 0 | 0 | 0 | 0 |
| methane metabolism | 0 | 0.19 | 0.118 | 0.14 | 0.138 |
| methionine metabolism | 0 | 0.175 | 0.048 | 0.164 | 0.177 |
| N -glycan biosynthesis | 0 | 0.2 | 0.301 | 0.329 | 0.369 |
| nictinate and nicotinamide metabolism | 0 | 0.636 | 0.667 | 0.758 | 0.746 |
| nitrogen metabolism | 0 | 0.075 | 0.115 | 0.081 | 0.082 |
| novobiocin biosynthesis | 0 | 0 | 0 | 0 | 0 |


| nucleotide sugars metabolism | 0 | 0.25 | 0.25 | 0.308 | 0.33 |
| :---: | :---: | :---: | :---: | :---: | :---: |
| one carbon pool by folate | 0 | 0.103 | 0.282 | 0.229 | 0.307 |
| oxidative phosphorylation | 0.372 | 0.632 | 0.825 | 0.831 | 0.852 |
| pantothenate and CoA biosynthesis | 0 | 0.24 | 0.264 | 0.301 | 0.31 |
| pentose and glucuronate interconversions | 0 | 0.118 | 0.182 | 0.289 | 0.337 |
| pentose phosphate pathway | 0 | 0.342 | 0.431 | 0.316 | 0.421 |
| peptidoglycan biosynthesis | 0 | 0 | 0 | 0 | 0 |
| phenylalanine metabolism | 0 | 0 | 0 | 0.019 | 0 |
| phenylalanine, tyrosine and tryptophan biosynthesis | 0 | 0.565 | 0.674 | 0.659 | 0.614 |
| phosphatidylinositol signaling system | 0 | 0.163 | 0.077 | 0.115 | 0.119 |
| polyunsaturated fatty acid biosynthesis | 0.667 | 0.703 | 0.828 | 0.906 | 0.813 |
| porphyrin and chlorophyll metabolism | 0 | 0.39 | 0.667 | 0.6 | 0.633 |
| proteasome | 0.143 | 0.782 | 0.931 | 0.952 | 0.949 |
| protein export | 0 | 0.846 | 0.909 | 0.984 | 0.966 |
| purine metabolism | 0.051 | 0.292 | 0.324 | 0.308 | 0.302 |
| pyrimidine metabolism | 0 | 0.186 | 0.295 | 0.311 | 0.312 |
| pyruvate metabolism | 0 | 0.164 | 0.227 | 0.233 | 0.205 |
| reductive carboxylate cycle (CO2 fixation) | 0 | 0 | 0 | 0.033 | 0.012 |
| regulation of autophagy | 0 | 0.533 | 0.866 | 0.915 | 0.863 |
| riboflavin metabolism | 0 | 0.313 | 0.64 | 0.581 | 0.6 |
| ribosome | 0.754 | 0.773 | 0.951 | 0.955 | 0.967 |
| RNA polymerase | 0 | 0.031 | 0.014 | 0.052 | 0.019 |
| selenoamino acid metabolism | 0 | 0.031 | 0.04 | 0.023 | 0.063 |
| SNARE interactions in vesicular transport | 0.3 | 0.784 | 0.893 | 0.956 | 0.95 |
| sphingolipid metabolism | 0 | 0.711 | 0.892 | 0.901 | 0.964 |
| starch and sucrose metabolism | 0.05 | 0.462 | 0.494 | 0.51 | 0.509 |
| streptomycin biosynthesis | 0 | 0 | 0 | 0.024 | 0.053 |
| styrene degradation | 0 | 0 | 0 | 0 | 0 |
| sulfur metabolism | 0 | 0.222 | 0.242 | 0.155 | 0.202 |
| synthesis and degradation of ketone bodies | 0 | 0 | 0 | 0.057 | 0.043 |
| taurine and hypotaurine metabolism | 0 | 0 | 0 | 0 | 0 |
| terpenoid biosynthesis | 0 | 0 | 0.231 | 0.157 | 0.163 |
| thiamine metabolism | 0 | 0.615 | 0.882 | 0.93 | 0.897 |
| tryptophan metabolism | 0 | 0.051 | 0.06 | 0.108 | 0.192 |
| two-component system | 0 | 0.333 | 0 | 0.182 | 0.056 |
| tyrosine metabolism | 0 | 0 | 0 | 0 | 0 |
| ubiquinone biosynthesis | 0 | 0.842 | 1 | 0.972 | 1 |
| ubiquitin mediated proteolysis | 0.296 | 0.508 | 0.617 | 0.604 | 0.609 |
| urea cycle and metabolism of amino groups | 0 | 0.393 | 0.449 | 0.458 | 0.399 |


| valine, leucine and isoleucine biosynthesis | 0.143 | 0.164 | 0.187 | 0.186 | 0.234 |
| :--- | :--- | ---: | ---: | :--- | :--- |
| valine, leucine and isoleucine degradation | 0 | 0 | 0 | 0.029 | 0 |
| vitamin B6 metabolism | 0 | 0.556 | 0.815 | 0.77 | 0.734 |

Table B1

## Appendix C: F-measures for the networks using the MultiBoostAB algorithm using the

 J48 classifierThe Table C1 below shows the F-measures for the networks using the MultiBoostAB algorithm using the J48 classifier trained with the properties of chromosome number, start position, stop position, strand, neighbor's number, neighbor's Gene Ontology term, neighbor's strand, neighbor's distance (determined from the mid-point of each feature), and network involvement information from KEGG for the genomic features (features being: genes, ARS, etc.). Measures for the ten-nearest neighbors could not be obtained.

| F-measures | Number of Neighbors |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Network | No | 1 | 2 | 5 | 10 |
| 1- and 2-methylnaphthalene degradation | 0 | 0.017 | 0.02 | 0 |  |
| 1,4-dichlorobenzene and dicarboxylate metabolism | 0 | 0 | 0.148 | 0 |  |
| 3-chloroacrylic acid degradation | 0 | 0.022 | 0 | 0.019 |  |
| ABC transporters | 0 | 0.7 | 1 | 1 |  |
| alanine and aspartate metabolism | 0 | 0.14 | 0.228 | 0.276 |  |
| alkaloid biosynthesis II | 0 | 0.118 | 0.307 | 0.323 |  |
| alpha-linolenic acid metabolism | 0 | 0.056 | 0.045 | 0.026 |  |
| aminoacyl-tRNA biosynthesis | 0 | 0.181 | 0.28 | 0.416 |  |
| aminophosphonate metabolism | 0 | 0.138 | 0.181 | 0.199 |  |
| aminosugars metabolism | 0 | 0.589 | 0.771 | 0.808 |  |
| arachidonic acid metabolism | 0 | 0.216 | 0.403 | 0.353 |  |
| arginine and proline metabolism | 0 | 0.218 | 0.369 | 0.423 |  |
| basal transcription factors | 0 | 0.77 | 1 | 1 |  |
| benzoate degradation via CoA ligation | 0 | 0.059 | 0.112 | 0.145 |  |
| benzoate degradation via hydroxylation | 0 | 0.067 | 0 | 0 |  |
| beta-alanine metabolism | 0 | 0.091 | 0.098 | 0.127 |  |
| bile acid biosynthesis | 0 | 0.03 | 0.118 | 0.142 |  |
| biosynthesis of steroids | 0 | 0.621 | 0.851 | 0.865 |  |
| biotin metabolism | 0 | 0.343 | 0.672 | 0.802 |  |
| butanoate metabolism | 0 | 0.088 | 0.072 | 0.146 |  |
| C5-branched dibasic acid metabolism | 0 | 0 | 0.098 | 0.302 |  |
| carbon fixation | 0 | 0.05 | 0.149 | 0.215 |  |


| cell communication | 0 | 1 | 1 | 1 |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| cell cycle | 0 | 0.578 | 0.858 | 0.876 |  |
| citrate cycle | 0 | 0.123 | 0.262 | 0.286 |  |
| cyanoamino acid metabolism | 0 | 0.061 | 0.092 | 0.124 |  |
| cysteine metabolism | 0 | 0.017 | 0.118 | 0.172 |  |
| DNA polymerase | 0 | 0.281 | 0.557 | 0.565 |  |
| ether lipid metabolism | 0 | 0.083 | 0.275 | 0.29 |  |
| ethylbenzene degradation | 0 | 0 | 0 | 0.092 |  |
| fatty acid biosynthesis | 0 | 0.235 | 0.556 | 0.619 |  |
| fatty acid elongation in mitochondria | 0 | 0.333 | 0.313 | 0 |  |
| fatty acid metabolism | 0 | 0.242 | 0.24 | 0.246 |  |
| fluorobenzoate degradation | 0 | 0 | 0.182 | 0 |  |
| folate biosynthesis | 0 | 0.29 | 0.491 | 0.54 |  |
| fructose and mannose metabolism | 0 | 0.263 | 0.373 | 0.353 |  |
| gamma-Hexachlorocyclohexane degradation | 0 | 0.139 | 0.415 | 0.35 |  |
| geraniol degradation | 0 | 0.111 | 0 | 0.091 |  |
| glutamate metabolism | 0 | 0.156 | 0.258 | 0.231 |  |
| glutathione metabolism | 0 | 0.217 | 0.304 | 0.38 |  |
| glycan structures biosynthesis 1 | 0 | 0.252 | 0.339 | 0.442 |  |
| glycan structures biosynthesis 2 | 0 | 0.187 | 0.34 | 0.42 |  |
| glycan structures degradation | 0 | 0.909 | 1 | 1 |  |
| glycerolipid metabolism | 0 | 0.268 | 0.378 | 0.435 |  |
| glycerophospholipid metabolism | 0 | 0.272 | 0.461 | 0.534 |  |
| glycine, serine and threonine metabolism | 0 | 0.25 | 0.315 | 0.35 |  |
| glycolysis/gluconeogenesis | 0 | 0.208 | 0.324 | 0.387 |  |
| glycosylphosphatidylinositol-anchor biosynthesis | 0 | 0.285 | 0.452 | 0.51 |  |
| glyoxylate and dicarboxylate metabolism | 0 | 0.211 | 0.326 | 0.4 |  |
| high-mannose type N -glycan biosynthesis | 0 | 0.336 | 0.405 | 0.51 |  |
| histidine metabolism | 0 | 0.283 | 0.38 | 0.394 |  |
| inositol phosphate metabolism | 0 | 0.062 | 0.251 | 0.274 |  |
| limonene and pinene degradation | 0 | 0.043 | 0.07 | 0.011 |  |
| lipoic acid metabolism | 0 | 0.69 | 1 | 1 |  |
| lysine biosynthesis | 0 | 0.32 | 0.559 | 0.603 |  |
| lysine degradation | 0 | 0.047 | 0.154 | 0.167 |  |
| MAPK signaling pathway | 0 | 0.576 | 0.807 | 0.826 |  |
| metabolism of xenobiotics by cytochrome P450 | 0 | 0.022 | 0.016 | 0.008 |  |
| methane metabolism | 0 | 0.113 | 0.138 | 0.166 |  |
| methionine metabolism | 0 | 0.187 | 0.262 | 0.343 |  |
| N -glycan biosynthesis | 0 | 0.346 | 0.517 | 0.54 |  |


| nictinate and nicotinamide metabolism | 0 | 0.514 | 0.761 | 0.806 |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| nitrogen metabolism | 0 | 0.116 | 0.181 | 0.313 |  |
| novobiocin biosynthesis | 0 | 0.067 | 0.023 | 0.083 |  |
| nucleotide sugars metabolism | 0 | 0.45 | 0.409 | 0.574 |  |
| one carbon pool by folate | 0 | 0.282 | 0.377 | 0.461 |  |
| oxidative phosphorylation | 0.361 | 0.633 | 0.908 | 0.924 |  |
| pantothenate and CoA biosynthesis | 0 | 0.248 | 0.429 | 0.554 |  |
| pentose and glucuronate interconversions | 0 | 0.231 | 0.4 | 0.375 |  |
| pentose phosphate pathway | 0 | 0.351 | 0.507 | 0.573 |  |
| peptidoglycan biosynthesis | 0 | 0.105 | 0.211 | 0.071 |  |
| phenylalanine metabolism | 0 | 0.053 | 0.023 | 0.113 |  |
| phenylalanine, tyrosine and tryptophan biosynthesis | 0 | 0.325 | 0.689 | 0.717 |  |
| phosphatidylinositol signaling system | 0 | 0.114 | 0.287 | 0.347 |  |
| polyunsaturated fatty acid biosynthesis | 0.4 | 0.762 | 0.915 | 0.924 |  |
| porphyrin and chlorophyll metabolism | 0 | 0.407 | 0.742 | 0.81 |  |
| proteasome | 0 | 0.667 | 1 | 1 |  |
| protein export | 0 | 0.699 | 1 | 1 |  |
| purine metabolism | 0 | 0.288 | 0.437 | 0.476 |  |
| pyrimidine metabolism | 0 | 0.243 | 0.417 | 0.452 |  |
| pyruvate metabolism | 0 | 0.126 | 0.308 | 0.365 |  |
| reductive carboxylate cycle (CO2 fixation) | 0 | 0.079 | 0.12 | 0.216 |  |
| regulation of autophagy | 0 | 0.612 | 0.944 | 0.967 |  |
| riboflavin metabolism | 0 | 0.275 | 0.744 | 0.805 |  |
| ribosome | 0.425 | 0.656 | 1 | 1 |  |
| RNA polymerase | 0 | 0.044 | 0.196 | 0.256 |  |
| selenoamino acid metabolism | 0 | 0.089 | 0.159 | 0.184 |  |
| SNARE interactions in vesicular transport | 0 | 0.658 | 0.999 | 1 |  |
| sphingolipid metabolism | 0 | 0.608 | 0.959 | 0.952 |  |
| starch and sucrose metabolism | 0 | 0.409 | 0.593 | 0.632 |  |
| streptomycin biosynthesis | 0 | 0.031 | 0.099 | 0.097 |  |
| styrene degradation | 0 | 0 | 0.087 | 0 |  |
| sulfur metabolism | 0 | 0.101 | 0.312 | 0.36 |  |
| synthesis and degradation of ketone bodies | 0 | 0.105 | 0.16 | 0.114 |  |
| taurine and hypotaurine metabolism | 0 | 0 | 0.059 | 0.089 |  |
| terpenoid biosynthesis | 0 | 0.349 | 0.391 | 0.467 |  |
| thiamine metabolism | 0 | 0.706 | 1 | 1 |  |
| tryptophan metabolism | 0 | 0.15 | 0.234 | 0.259 |  |
| two-component system | 0 | 0.316 | 0.44 | 0.489 |  |
| tyrosine metabolism | 0 | 0.019 | 0.031 | 0.058 |  |


| ubiquinone biosynthesis | 0 | 0.41 | 1 | 1 |  |
| :--- | ---: | ---: | ---: | ---: | ---: |
| ubiquitin mediated proteolysis | 0 | 0.451 | 0.705 | 0.752 |  |
| urea cycle and metabolism of amino groups | 0 | 0.279 | 0.52 | 0.575 |  |
| valine, leucine and isoleucine biosynthesis | 0 | 0.278 | 0.304 | 0.351 |  |
| valine, leucine and isoleucine degradation | 0 | 0.028 | 0.052 | 0.091 |  |
| vitamin B6 metabolism | 0 | 0.5 | 0.82 | 0.87 |  |

Table C1

## Appendix D: F-measures for the networks using KStar classifier

The Table D1 shows the F-measures for the networks using the KStar classifier trained with the properties of chromosome number, start position, stop position, strand, neighbor's number, neighbor's network involvement information from KEGG, neighbor's strand, neighbor's distance (determined from the mid-point of each feature), and network involvement information from KEGG for genomic features (features being: genes, ARS, etc.).

| F-measure | K* |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Network | No | 1 | 2 | 5 | 10 |
| 1- and 2-methylnaphthalene degradation | 0 | 0.297 | 0.277 | 0.297 | 0.275 |
| 1,4-dichlorobenzene and dicarboxylate metabolism | 0 | 0.118 | 0.118 | 0.156 | 0.171 |
| 3-chloroacrylic acid degradation | 0 | 0.355 | 0.417 | 0.474 | 0.436 |
| ABC transporters | 0 | 0.556 | 0.267 | 0.333 | 0.327 |
| alanine and aspartate metabolism | 0 | 0.212 | 0.301 | 0.312 | 0.319 |
| alkaloid biosynthesis II | 0 | 0.571 | 0.483 | 0.387 | 0.387 |
| alpha-linolenic acid metabolism | 0 | 0 | 0.364 | 0.25 | 0.069 |
| aminoacyl-tRNA biosynthesis | 0.069 | 0.222 | 0.247 | 0.211 | 0.222 |
| aminophosphonate metabolism | 0 | 0 | 0 | 0 | 0.063 |
| aminosugars metabolism | 0 | 0.821 | 0.832 | 0.796 | 0.792 |
| arachidonic acid metabolism | 0 | 0.101 | 0.107 | 0.088 | 0.113 |
| arginine and proline metabolism | 0 | 1 | 1 | 1 | 0.995 |
| basal transcription factors | 0.2 | 0.863 | 0.878 | 0.878 | 0.856 |
| benzoate degradation via CoA ligation | 0 | 0.356 | 0.519 | 0.502 | 0.427 |
| benzoate degradation via hydroxylation | 0 | 0.313 | 0.319 | 0.351 | 0.321 |
| beta-alanine metabolism | 0 | 0.335 | 0.387 | 0.305 | 0.301 |
| bile acid biosynthesis | 0 | 0.116 | 0.138 | 0.097 | 0.145 |
| biosynthesis of steroids | 0 | 0.135 | 0.108 | 0.131 | 0.153 |
| biotin metabolism | 0 | 0.137 | 0.22 | 0.215 | 0.179 |
| butanoate metabolism | 0 | 0.074 | 0.167 | 0.162 | 0.173 |
| C5-branched dibasic acid metabolism | 0 | 0 | 0.067 | 0.014 | 0.007 |
| carbon fixation | 0 | 0 | 0.038 | 0 | 0.009 |
| cell communication | 0 | 0 | 0 | 0 | 0 |
| cell cycle | 0.123 | 0.273 | 0.186 | 0.234 | 0.205 |
| citrate cycle | 0 | 0.375 | 0.414 | 0.363 | 0.389 |


| cyanoamino acid metabolism | 0 | 0.11 | 0.088 | 0.111 | 0.079 |
| :---: | :---: | :---: | :---: | :---: | :---: |
| cysteine metabolism | 0 | 0.35 | 0.277 | 0.286 | 0.294 |
| DNA polymerase | 0 | 0.359 | 0.446 | 0.373 | 0.366 |
| ether lipid metabolism | 0 | 0 | 0 | 0 | 0.003 |
| ethylbenzene degradation | 0 | 0 | 0 | 0 | 0 |
| fatty acid biosynthesis | 0 | 0.324 | 0.219 | 0.212 | 0.203 |
| fatty acid elongation in mitochondria | 0 | 0 | 0 | 0 | 0 |
| fatty acid metabolism | 0 | 0 | 0 | 0 | 0 |
| fluorobenzoate degradation | 0 | 0.107 | 0.217 | 0.163 | 0.16 |
| folate biosynthesis | 0 | 0.679 | 0.524 | 0.583 | 0.626 |
| fructose and mannose metabolism | 0 | 0 | 0 | 0 | 0 |
| gamma-Hexachlorocyclohexane degradation | 0 | 0.077 | 0.087 | 0.024 | 0.038 |
| geraniol degradation | 0 | 0 | 0 | 0.069 | 0 |
| glutamate metabolism | 0 | 0 | 0.048 | 0.063 | 0.058 |
| glutathione metabolism | 0 | 0 | 0.042 | 0.038 | 0.038 |
| glycan structures biosynthesis 1 | 0 | 0 | 0 | 0.068 | 0.022 |
| glycan structures biosynthesis 2 | 0 | 0 | 0.197 | 0.14 | 0.143 |
| glycan structures degradation | 0 | 0.497 | 0.534 | 0.501 | 0.533 |
| glycerolipid metabolism | 0 | 0.452 | 0.307 | 0.394 | 0.367 |
| glycerophospholipid metabolism | 0.125 | 0.348 | 0.226 | 0.248 | 0.266 |
| glycine, serine and threonine metabolism | 0 | 0.182 | 0.32 | 0.291 | 0.333 |
| glycolysis/gluconeogenesis | 0 | 0 | 0 | 0.031 | 0.01 |
| glycosylphosphatidylinositol-anchor biosynthesis | 0 | 0.745 | 0.643 | 0.684 | 0.684 |
| glyoxylate and dicarboxylate metabolism | 0 | 0 | 0 | 0 | 0 |
| high-mannose type N -glycan biosynthesis | 0 | 0.373 | 0.333 | 0.304 | 0.249 |
| histidine metabolism | 0 | 0.212 | 0.073 | 0.173 | 0.119 |
| inositol phosphate metabolism | 0 | 0.226 | 0.304 | 0.327 | 0.256 |
| limonene and pinene degradation | 0 | 0.377 | 0.355 | 0.328 | 0.411 |
| lipoic acid metabolism | 0 | 0 | 0.067 | 0.118 | 0.103 |
| lysine biosynthesis | 0 | 0.3 | 0.2 | 0.293 | 0.273 |
| lysine degradation | 0 | 0 | 0 | 0 | 0 |
| MAPK signaling pathway | 0 | 0.944 | 0.966 | 0.933 | 0.918 |
| metabolism of xenobiotics by cytochrome P450 | 0 | 0.282 | 0.196 | 0.232 | 0.22 |
| methane metabolism | 0 | 0 | 0 | 0 | 0 |
| methionine metabolism | 0 | 0 | 0.222 | 0 | 0 |
| N -glycan biosynthesis | 0 | 0.182 | 0.175 | 0.231 | 0.187 |
| nictinate and nicotinamide metabolism | 0 | 0 | 0.081 | 0.054 | 0.026 |
| nitrogen metabolism | 0 | 0 | 0 | 0 | 0 |
| novobiocin biosynthesis | 0 | 0 | 0 | 0 | 0 |


| nucleotide sugars metabolism | 0 | 0 | 0 | 0 | 0.08 |
| :---: | :---: | :---: | :---: | :---: | :---: |
| one carbon pool by folate | 0 | 0.027 | 0.015 | 0.029 | 0.028 |
| oxidative phosphorylation | 0.372 | 0 | 0 | 0 | 0.027 |
| pantothenate and CoA biosynthesis | 0 | 0.25 | 0.158 | 0.22 | 0.261 |
| pentose and glucuronate interconversions | 0 | 0 | 0.114 | 0.15 | 0.108 |
| pentose phosphate pathway | 0 | 0 | 0.053 | 0.009 | 0.064 |
| peptidoglycan biosynthesis | 0 | 0.067 | 0.034 | 0.034 | 0.054 |
| phenylalanine metabolism | 0 | 1 | 1 | 1 | 0.936 |
| phenylalanine, tyrosine and tryptophan biosynthesis | 0 | 0.424 | 0.579 | 0.673 | 0.613 |
| phosphatidylinositol signaling system | 0 | 0.8 | 0.873 | 0.85 | 0.74 |
| polyunsaturated fatty acid biosynthesis | 0.667 | 0.9 | 0.821 | 0.77 | 0.784 |
| porphyrin and chlorophyll metabolism | 0 | 0.4 | 0.393 | 0.408 | 0.319 |
| proteasome | 0.143 | 0.714 | 0.75 | 0.679 | 0.683 |
| protein export | 0 | 1 | 1 | 1 | 0.97 |
| purine metabolism | 0.051 | 0.273 | 0.319 | 0.369 | 0.364 |
| pyrimidine metabolism | 0 | 0.607 | 0.76 | 0.732 | 0.668 |
| pyruvate metabolism | 0 | 0.167 | 0.16 | 0.231 | 0.163 |
| reductive carboxylate cycle (CO2 fixation) | 0 | 0 | 0 | 0.027 | 0.02 |
| regulation of autophagy | 0 | 0.085 | 0.076 | 0.08 | 0.106 |
| riboflavin metabolism | 0 | 0.167 | 0.185 | 0.26 | 0.299 |
| ribosome | 0.754 | 0.143 | 0.171 | 0.11 | 0.083 |
| RNA polymerase | 0 | 0.14 | 0.147 | 0.15 | 0.172 |
| selenoamino acid metabolism | 0 | 0 | 0 | 0 | 0.008 |
| SNARE interactions in vesicular transport | 0.3 | 0.158 | 0.188 | 0.141 | 0.192 |
| sphingolipid metabolism | 0 | 0.261 | 0.239 | 0.189 | 0.228 |
| starch and sucrose metabolism | 0.05 | 1 | 1 | 1 | 1 |
| streptomycin biosynthesis | 0 | 0.885 | 0.895 | 0.833 | 0.866 |
| styrene degradation | 0 | 1 | 1 | 1 | 0.706 |
| sulfur metabolism | 0 | 1 | 1 | 1 | 0.963 |
| synthesis and degradation of ketone bodies | 0 | 0.25 | 0 | 0.25 | 0.092 |
| taurine and hypotaurine metabolism | 0 | 0.968 | 1 | 1 | 0.975 |
| terpenoid biosynthesis | 0 | 0.026 | 0.015 | 0.042 | 0.052 |
| thiamine metabolism | 0 | 0.92 | 1 | 1 | 0.957 |
| tryptophan metabolism | 0 | 0.34 | 0.512 | 0.559 | 0.56 |
| two-component system | 0 | 0.938 | 1 | 1 | 0.97 |
| tyrosine metabolism | 0 | 1 | 1 | 1 | 0.985 |
| ubiquinone biosynthesis | 0 | 0.758 | 0.777 | 0.791 | 0.768 |
| ubiquitin mediated proteolysis | 0.296 | 0.115 | 0.104 | 0.108 | 0.172 |
| urea cycle and metabolism of amino groups | 0 | 0.796 | 0.836 | 0.829 | 0.795 |


| valine, leucine and isoleucine biosynthesis | 0.143 | 0.6 | 0.655 | 0.658 | 0.621 |
| :--- | :--- | ---: | ---: | ---: | ---: |
| valine, leucine and isoleucine degradation | 0 | 1 | 1 | 1 | 0.979 |
| vitamin B6 metabolism | 0 | 0.933 | 0.964 | 0.948 | 0.883 |

Table D1

## Appendix E: F-measures for the Gene Ontology term using the MultiBoostAB algorithm

## using the J48 classifier

The Table E1 below shows the F-measures for the networks using the MultiBoostAB algorithm using the J48 classifier trained with the properties of chromosome number, start position, stop position, strand, neighbor's number, neighbor's SGDID, neighbor's strand, neighbor's distance (determined from the mid-point of each feature), and Gene Ontology information for the genomic features (features being: genes, ARS, etc.).

| F-measures | Number of Neighbors |  |  |  |
| :--- | ---: | ---: | ---: | ---: |
| GO terms | No | $\mathbf{1}$ | $\mathbf{2}$ | $\mathbf{5}$ |
| amino acid and derivative metabolic process | 0 | 0 | 0 |  |
| anatomical structure morphogenesis | 0 | 0 | 0 |  |
| biological process | 0.368 | 0.4 | 0.403 |  |
| carbohydrate metabolic process | 0 | 0 | 0 |  |
| cell budding | 0 | 0 | 0 |  |
| cell cortex | 0 | 0 | 0 |  |
| cell cycle | 0 | 0 | 0 |  |
| cell wall | 0 | 0 | 0 |  |
| cell wall organization and biogenesis | 0 | 0 | 0 | 0 |
| cellular bud | 0 | 0 | 0 |  |
| cellular component | 0 | 0 | 0 |  |
| cellular homeostasis | 0.191 | 0 | 0 | 0 |
| cellular respiration | 0 | 0 | 0 |  |
| chromosome | 0 | 0 | 0 | 0 |
| conjugation | 0 | 0 | 0 |  |
| cytokinesis | 0 | 0 | 0 |  |
| cytoplasm | 0 | 0 | 0.358 | 0.363 |
| cytoplasmic membrane-bound vesicle | 0.25 | 0 | 0 |  |
| cytoskeleton | 0 | 0 | 0 | 0 |
| cytoskeleton organization and biogenesis | 0 | 0 | 0 | 0 |
| DNA binding | 0 | 0 | 0 | 0 |
| DNA metabolic process | 0 | 0 |  |  |
| electron transport | 0.054 | 0 | 0 |  |
| endomembrane system | 0 | 0 | 0 |  |
| endoplasmic reticulum | 0 | 0 | 0 |  |
| enzyme regulator activity | 0 | 0 | 0 |  |
| extracellular region | 0 | 0 | 0 | 0 |
| generation of precursor metabolites and energy | 0 | 0 | 0 | 0 |
| Golgi apparatus | 0 | 0 | 0 | 0 |
| helicase activity | 0 | 0 | 0 | 0 |
| hydrolase activity | 0 | 0 | 0 |  |
| isomerase activity | 0 | 0 | 0 |  |
| ligase activity | 0 | 0 | 0 | 0 |


| lipid metabolic process | 0 | 0 | 0 |  |
| :---: | :---: | :---: | :---: | :---: |
| lyase activity | 0 | 0 | 0 |  |
| meiosis | 0 | 0 | 0 |  |
| membrane | 0.027 | 0 | 0 |  |
| membrane fraction | 0 | 0 | 0 |  |
| membrane organization and biogenesis | 0 | 0 | 0 |  |
| microtubule organizing center | 0 | 0 | 0 |  |
| mitochondrial envelope | 0 | 0 | 0 |  |
| mitochondrion | 0.17 | 0 | 0 |  |
| molecular function | 0.572 | 0.584 | 0.574 |  |
| motor activity | 0 | 0 | 0 |  |
| nuclear organization and biogenesis | 0.154 | 0 | 0 |  |
| nucleolus | 0.112 | 0 | 0 |  |
| nucleotidyltransferase activity | 0 | 0 | 0 |  |
| nucleus | 0.188 | 0 | 0 |  |
| organelle organization and biogenesis | 0.022 | 0 | 0 |  |
| oxidoreductase activity | 0 | 0 | 0 |  |
| peptidase activity | 0 | 0 | 0 |  |
| peroxisome | 0 | 0 | 0 |  |
| phosphoprotein phosphatase activity | 0 | 0 | 0 |  |
| plasma membrane | 0.014 | 0 | 0 |  |
| protein binding | 0 | 0 | 0 |  |
| protein catabolic process | 0.042 | 0 | 0 |  |
| protein kinase activity | 0 | 0 | 0 |  |
| protein modification process | 0.026 | 0 | 0 |  |
| pseudohyphal growth | 0 | 0 | 0 |  |
| response to stress | 0.016 | 0 | 0 |  |
| ribosome | 0.028 | 0 | 0 |  |
| ribosome biogenesis and assembly | 0.011 | 0 | 0 |  |
| RNA binding | 0 | 0 | 0 |  |
| RNA metabolic process | 0.061 | 0 | 0 |  |
| signal transducer activity | 0 | 0 | 0 |  |
| signal transduction | 0 | 0 | 0 |  |
| site of polarized growth | 0 | 0 | 0 |  |
| sporulation | 0 | 0 | 0 |  |
| structural molecule activity | 0 | 0 | 0 |  |
| transcription | 0.054 | 0 | 0 |  |
| transcription regulator activity | 0 | 0 | 0 |  |
| transferase activity | 0 | 0 | 0 |  |
| translation | 0.122 | 0 | 0 |  |
| translation regulator activity | 0 | 0 | 0 |  |
| transport | 0.074 | 0 | 0 |  |
| transporter activity | 0 | 0 | 0 |  |
| vacuole | 0 | 0 | 0 |  |
| vesicle-mediated transport | 0.011 | 0 | 0 |  |
| vitamin metabolic process | 0.083 | 0 | 0 |  |

## Table E1

## Appendix F: F-measures for the Gene Ontology term using the MultiBoostAB algorithm

 using the J48 classifierThe Table F1 below shows the F-measures for the networks using the MultiBoostAB algorithm using the J48 classifier trained with the properties of chromosome number, start position, stop position, strand, neighbor's number, neighbor's Gene Ontology information, neighbor's strand, neighbor's distance (determined from the mid-point of each feature), and Gene Ontology information for the genomic features (features being: genes, ARS, etc.).

| F-measures | Number of Neighbors |  |  |  |
| :--- | ---: | ---: | ---: | ---: |
| GO terms | No | $\mathbf{1}$ |  |  |
| amino acid and derivative metabolic process | 0 | $\mathbf{2}$ | $\mathbf{5}$ |  |
| anatomical structure morphogenesis | 0 | 0.838 | 0.833 |  |
| biological process | 0 | 0.333 | 0.41 |  |
| carbohydrate metabolic process | 0.368 | 0.4 | 0.949 | 0.954 |
| cell budding | 0 | 0 | 0.8 | 0.806 |
| cell cortex | 0 | 0 | 0.757 | 0.758 |
| cell cycle | 0 | 0 | 0.373 | 0.452 |
| cell wall | 0 | 0 | 0.64 | 0.631 |
| cell wall organization and biogenesis | 0 | 0.025 | 0.96 | 0.993 |
| cellular bud | 0 | 0 | 0.788 | 0.78 |
| cellular component | 0 | 0 | 0.365 | 0.411 |
| cellular homeostasis | 0.191 | 0.307 | 0.861 | 0.885 |
| cellular respiration | 0 | 0 | 0.805 | 0.742 |
| chromosome | 0 | 0 | 0.779 | 0.76 |
| conjugation | 0 | 0.057 | 0.556 | 0.594 |
| cytokinesis | 0 | 0 | 0.727 | 0.72 |
| cytoplasm | 0 | 0 | 0.843 | 0.862 |
| cytoplasmic membrane-bound vesicle | 0.25 | 0.364 | 0.702 | 0.717 |
| cytoskeleton | 0 | 0.049 | 0.778 | 0.777 |
| cytoskeleton organization and biogenesis | 0 | 0.035 | 0.521 | 0.532 |
| DNA binding | 0 | 0 | 0.724 | 0.738 |
| DNA metabolic process | 0 | 0 | 0.756 | 0.789 |
| electron transport | 0.054 | 0 | 0.762 | 0.758 |
| endomembrane system | 0 | 0 | 0.795 | 0.812 |
| endoplasmic reticulum | 0 | 0.023 | 0.539 | 0.493 |
| enzyme regulator activity | 0.026 | 0.116 | 0.65 | 0.659 |
| extracellular region | 0 | 0 | 0.915 | 0.952 |
| generation of precursor metabolites and energy | 0 | 0.074 | 0.741 | 0.808 |
| Golgi apparatus | 0 | 0 | 0.869 | 0.867 |
| helicase activity | 0 | 0.029 | 0.766 | 0.807 |
| hydrolase activity | 0 | 0 | 0.882 | 0.921 |
| isomerase activity | 0 | 0 | 0.93 | 0.951 |
| ligase activity | 0 | 0 | 0.849 | 0.91 |
|  | 0 | 0 | 0.981 | 0.985 |


| lipid metabolic process | 0 | 0 | 0.893 | 0.889 |
| :---: | :---: | :---: | :---: | :---: |
| lyase activity | 0 | 0 | 0.917 | 0.947 |
| meiosis | 0 | 0 | 0.66 | 0.624 |
| membrane | 0.027 | 0.086 | 0.5 | 0.519 |
| membrane fraction | 0 | 0.14 | 0.635 | 0.649 |
| membrane organization and biogenesis | 0 | 0 | 0.589 | 0.582 |
| microtubule organizing center | 0 | 0 | 0.741 | 0.762 |
| mitochondrial envelope | 0 | 0 | 0.668 | 0.682 |
| mitochondrion | 0.17 | 0.188 | 0.774 | 0.801 |
| molecular function | 0.572 | 0.572 | 0.951 | 0.987 |
| motor activity | 0 | 0 | 0.928 | 0.93 |
| nuclear organization and biogenesis | 0.154 | 0 | 0.752 | 0.733 |
| nucleolus | 0.112 | 0.063 | 0.758 | 0.799 |
| nucleotidyltransferase activity | 0 | 0 | 0.931 | 0.993 |
| nucleus | 0.188 | 0.3 | 0.594 | 0.622 |
| organelle organization and biogenesis | 0.022 | 0 | 0.695 | 0.677 |
| oxidoreductase activity | 0 | 0 | 0.912 | 0.937 |
| peptidase activity | 0 | 0 | 0.958 | 0.963 |
| peroxisome | 0 | 0 | 0.735 | 0.805 |
| phosphoprotein phosphatase activity | 0 | 0 | 0.962 | 0.99 |
| plasma membrane | 0.014 | 0.076 | 0.714 | 0.726 |
| protein binding | 0 | 0 | 0.923 | 0.96 |
| protein catabolic process | 0.042 | 0 | 0.905 | 0.913 |
| protein kinase activity | 0 | 0 | 0.894 | 0.933 |
| protein modification process | 0.026 | 0 | 0.796 | 0.806 |
| pseudohyphal growth | 0 | 0 | 0.883 | 0.889 |
| response to stress | 0.016 | 0 | 0.871 | 0.874 |
| ribosome | 0.028 | 0.074 | 0.726 | 0.767 |
| ribosome biogenesis and assembly | 0.011 | 0 | 0.837 | 0.832 |
| RNA binding | 0 | 0 | 0.933 | 0.971 |
| RNA metabolic process | 0.061 | 0 | 0.747 | 0.746 |
| signal transducer activity | 0 | 0 | 0.813 | 0.835 |
| signal transduction | 0 | 0 | 0.707 | 0.721 |
| site of polarized growth | 0 | 0.032 | 0.43 | 0.428 |
| sporulation | 0 | 0 | 0.59 | 0.625 |
| structural molecule activity | 0 | 0 | 0.95 | 0.982 |
| transcription | 0.054 | 0 | 0.829 | 0.829 |
| transcription regulator activity | 0 | 0 | 0.827 | 0.875 |
| transferase activity | 0 | 0 | 0.954 | 0.987 |
| translation | 0.122 | 0 | 0.973 | 0.975 |
| translation regulator activity | 0 | 0 | 0.879 | 0.912 |
| transport | 0.074 | 0 | 0.756 | 0.748 |
| transporter activity | 0 | 0 | 0.894 | 0.934 |
| vacuole | 0 | 0.04 | 0.631 | 0.679 |
| vesicle-mediated transport | 0.011 | 0 | 0.691 | 0.7 |
| vitamin metabolic process | 0.083 | 0 | 0.847 | 0.855 |

## Table F1.

## Appendix G: NNge Rule Adaption

The Nearest-neighbor-like algorithm uses non-nested generalized exemplars which are hyperrectangles. These hyperrectangles can be viewed as if-then rules. An example of one of these rules follows:

```
class '(306406.2-459592.3]' IF : strand in {C} ^ network in
{sce03022} ^ chromosome in {sixteen}
```

This rule was adapted to:

If the feature is in network sce 03022 then look at strand C of chromosome sixteen in the range of 306406-459592.

There were several rules that mentioned this network, so at the end of the adaption process, there were several locations that were recommended. Below is the rule generated for this network, it is written in Java:

```
if(network.equals("sce03022"))
{
    addLocation("Chromosome two, Strand C, range: 612778 -
    765964");
    addLocation("Chromosome three, Strand C, range: 153220 -
    306406");
    addLocation("Chromosome four, Strand W, range: 1072336 -
    1225522");
    addLocation("Chromosome five, Strand W or C, range: 459592 -
    612778");
    addLocation("Chromosome seven, Strand W, range: 765964 -
    919150");
    addLocation("Chromosome seven, Strand C, range: 919150 -
    1072336");
    addLocation("Chromosome eleven, Strand W, range: 459592 -
    612778");
    addLocation("Chromosome twelve, Strand W, range: 153220 -
    306406");
    addLocation("Chromosome thirteen, Strand W or C, range: 612778
    - 765964");
    addLocation("Chromosome fifteen, Strand C, range: 612778 -
    765964");
```

```
    addLocation("Chromosome sixteen, Strand C, range: 306406 -
459592");
addLocation("Chromosome sixteen, Strand W, range: 612778 -
765964");
}
```


## Appendix H. Decision Trees Generated for Classifying Features to Networks

In the tree structure, a colon introduces the name of the network that has been assigned to a particular leaf. The network name is followed by the number of instances that reach that leaf. The number is expressed as a decimal because of the way the classification algorithm uses fractions to handle missing values. The number of incorrectly classified instances that reach that leaf, if they exist, are reported after the "/". For example, if the numbers $(2.0 / 1.0)$ are reported, it means that two instances reached the leaf, of which one is incorrectly classified.

## Decision Tree Generated for Chromosome One:

```
start <= 71787
| strand = C
| | start <= 45023: sce00720 (58.75/38.43)
| | start > 45023
| | | start <= 67521: sce04111 (9.31)
| | | start > 67521: sce00860 (9.85)
| strand = W
| | start <= 62841
| | | start <= 45023: sce00251 (21.99/9.33)
| | | start > 45023: sce04010 (10.77)
| | start > 62841: sce00010 (63.23/42.98)
start > 71787
| strand = C
| | start <= 105873: sce04111 (13.56)
| | start > 105873
| | | start <= 130802: sce04130 (11.84)
| | | start > 130802: sce04111 (8.9)
| strand = W
| | start <= 130802
| | | start <= 105873: sce04130 (7.54)
| | | start > 105873: sce00910 (106.76/82.31)
| | start > 130802
| | | start <= 175133: sce00230 (10.85)
| | | start > 175133
| | | | start <= 190187: sce00252 (9.53)
| | | start > 190187: sce00361 (8.11/3.17)
```


## Decision Tree Generated for Chromosome Two:

```
start <= 505662
| start <= 245669
| | start <= 168426
| | | start <= 80728
| | | | start <= 69442
```












## Decision Tree Generated for Chromosome Three:

```
start <= 105548
| start <= 68333
| | strand = C
| | | start <= 35393
| | | | start <= 16880
    | neigh_num <= 1: sce00260 (16.25/7.72)
    | neigh_num > 1: sce00272 (12.13/5.29)
    start > 16}880: sce01031 (30.85/8.96)
    start > 35393
        | start <= 50838: sce00920 (39.89/18.77)
        start > 50838: sce00340 (17.93)
        strand = W
            start <= 38801: sce00450 (65.19/49.58)
            start > 38801
            neigh_GO_term = membrane_: sce00521 (5.95/2.38)
            neigh_GO_term = transporter_activity_: sce00010
(0.0)
| | | | neigh_GO_term = transport_: sce00010 (0.0)
                    neigh_GO_term = vacuole_:- sce00010 (12.03/3.36)
            neigh_GO_term = cellular_component_: sce00010 (0.0)
            neigh_GO_term = molecular_function_
            | neigh_num <= 1: sce00}
            neigh_num > 1
            | distance <= 994: sce00500 (4.57/1.2)
                        distance > 994: sce00521 (5.7/2.64)
                neigh_GO_term = biological_process_: sce00010
(11.14/5.23)
| | | | neigh_GO_term = nucleus_: sce00500 (8.61/2.79)
| | | | neigh_GO_term = transcription_regulator_activity_:
sce00010 (0.0)
```






```
| | | | neigh GO term = signal transducer activity :
sce00051 (0.0)
| | | | neigh_GO_term =
phosphoprotein_phosphatase_activity_: sce00051 (0.0)
| | | | neigh_GO_term =
cytoskeleton_organization_and_biogenesis_: sce00051 (0.0)
```


## Decision Tree Generated for Chromosome Four:

```
start <= 931125
| start <= 159605
| | start <= 38488
| | | strand = W
    | start <= 33918: sce00361 (14.28)
    | start > 33918: sce00510 (32.4/11.71)
    strand = C
        | start <= 18566
        start <= 9756: sce00051 (10.32)
        start > 9756
            | neigh_strand = W
            | nēigh_num <= 1: sce00051 (12.49/6.41)
            | neigh_num > 1: sce00310 (17.12/9.46)
            neigh_strand = C: sce00260 (30.17/22.51)
        start > 18566
        | start <= 30454: sce00230 (17.06)
        start > 30454: sce02021 (46.75/16.32)
        start > 38488
        start <= 155641
        start <= 130485
        start <= 93746
            | start <= 73919: sce00910 (28.04/12.22)
            start > 73919: sce00860 (10.02)
            start > 93746
            | strand = W
                    | start <= 121593: sce03010 (16.79)
                        start > 121593: sce00190 (14.47)
                    strand = C
            | start <= 121593: sce04120 (13.19)
            | start > 121593: sce03010 (14.44)
        start > 130485
        start <= 147590: sce00620 (46.64/14.87)
        start > 147590: sce00251 (33.37/16.25)
            start > 155641
                neigh_GO_term = mitochondrion_: sce00120 (0.0)
                neigh_GO_term = signal_transducer_activity_:
sce00120 (0.0)
                        neigh_GO_term = biological_process
                        nēig\overline{h}}\mathrm{ strand = W: sce000071 (10.022/7.02)
                        neigh strand = C: sce00010 (12.41/9.4)
                        neigh_GO_term = cellular_component_: sce00641
(6.45/4.94)
```




| | | | | | | | | neigh GO_term = enzyme_regulator_activity_: sce04111 (0.0) | | | | | | | | | neigh_GO_term = cell_wall_organization_and_biogenesis_: sce04111 (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad$ neigh_GO_term = signal_transduction_: sce04111 (0.0)
| | | | | - | | | neigh GO_term = membrane : sce04111 (0.0)
| | | | | | | | | neigh_GO_term =
protein_binding_: sce04111 (0.0)
$\mid$ | | | | | | | neigh_GO_term = sporulation_:
sce04111 (0.0)
| | | | | | | | neigh_GO_term = hydrolase_activity_: sce04120 (7.79/0.86)
| | | | | | | | | neigh_GO_term = transferase_activity_: sce04111 (0.45/0.15)
| | | | | | | | | neigh_GO_term = response_to_stress_: sce04111 (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ neigh_GO_term = nucleus_:
sce04120 (7.9/3.81)
| | | | | | | | 1 neigh_GO_term = protein_modification_process_: sce04111 ( $\overline{0} .0 \overline{)}$
$|\quad| \quad|\quad| \quad|\quad| \quad$ neigh GO term = vesiclemediated_transport_: sce04111 (1.03/0.34)
| | | | | | | | neigh_GO_term =
endomembrane_system_: sce04111 (0.0)
| | | | | | | | | neigh_GO_term = endoplasmic_reticulum_: sce04111 (0.0)
| | | | | | | | neigh GO term = phosphoprotein_phosphatase_activity_: sce 04111 (0.0)
| | | | | | | | neigh_GO_term = translation_: sce04120 (3.3/0.6)
| | | | | | | | | neigh_GO_term = cell_cortex_:
sce04111 (0.0)
| | | | | | | | neigh_GO_term = cytoskeleton_:
sce04111 (0.0)
| | | | | | | | | neigh_GO_term =
structural_molecule_activity_: sce04111 ( $\overline{1} .7 \overline{1} / 0.31$ )
| | | | | | | $|\quad|$ neigh_GO_term = cytokinesis_: sce04111 (0.0)
| | | | | | | | neigh_GO_term =
DNA_metabolic_process_: sce04111 (0.0)
| $\left.|\quad| \quad\right|^{-}| |^{-}|\quad| \quad$ neigh_GO_term = cell_cycle_:
sce04111 (5.39/0.73)
| | | | | | | | | neigh_GO_term = RNA_binding_:
sce04111 (0.0)
| | | | | | | | neigh_GO_term =
anatomical_structure_morphogenesis_: sce04111 (0.0)
$|||||||\mid$ neigh_GO_term $=$ conjugation_:
sce04111 (0.0)
| | | | | | | | | neigh_GO_term = chromosome_:
sce04111 (0.0)
| | | | | | | | | neigh_GO_term = DNA_binding_:
sce04111 (0.0)

```
| | | | | | | | | neigh GO term =
membrane_organization_and_biogenesis_: sce\overline{04}\overline{1}11 (0.0)
| | | | | | | | | neigh_GO_term =
peptidase_activity_: sce04111 (0.0)
| | | | | | | | | neigh_GO_term =
protein kinase activity : sce04111 (0.0)
| | - | | | | - | | | neigh_GO_term = nucleolus_:
sce04111 (0.0)
| | | | | | | | | neigh_GO_term =
ribosome_biogenesis_and_assembly_: sce04111 (0.0)
| | | | | | | | | neigh_GO_term = vacuole_:
sce04111 (0.0)
| | | | | | | | | neigh_GO_term =
RNA_metabolic_process_: sce04111 (0.0)
| | | | | | | | | neigh_GO_term =
organelle_organization_and_biogenesis_: sce04111 (0.0)
| | | | | | | | | nēigh_GO_term = transcription_:
sce04111 (0.0)
| | | | | | | | | neigh_GO_term =
Golgi_apparatus_: sce04111 (0.0)
M
cytoplasmic_membrane-bound_vesicle_: sce04111 (0.0)
| | | | | | | | | neigh_GO_term = ribosome_:
sce04120 (1.67/0.67)
| | | | | | | | | neigh_GO_term =
amino_acid_and_derivative_metabolic_process_: sce04111 (4.97/0.99)
| | | | | | | | | neigh_GO_term =
generation_of_precursor_metabolites_and_energy_: sce04111 (0.0)
| | | | | | | | | neiğh GO tērm =
nucleotidyltransferase_activity_: sce0411\overline{1}
| | | | | | | | | neigh_GO_term =
transcription_regulator_activity_: sce041\overline{11}
| | | | | | | | | neigh_GO_term =
ligase_activity_: sce04111 (0.0)
| | | | | | | | | neigh_GO_term =
pseudohyphal_growth_: sce04111 (0.0)
| | | | | | | | | neigh_GO_term = meiosis_:
sce04111 (0.0)
| | | | | | | | | neigh_GO_term =
membrane_fraction_: sce04111 (0.0)
| | \ | | | | | | neigh_GO_term =
protein_catabolic_process_: sce04111 (0.0)
| | | | | | | | | neigh_GO_term = cellular_bud_:
sce04111 (0.0)
| | | | | | | | | neigh_GO_term =
cytoskeleton_organization_and_biogenesis_: sce04111 (0.0)
| | | | | | | | | neigh_GO_term =
mitochondrial_envelope_: sce04111 (0.0)
| | | | | | | | | neigh_GO_term =
site_of_polarized_growth_: sce04111 (0.0)
| | | | | | | | | neigh_GO_term =
lipid_metabolic_process_: sce04111 (0.0)
| | | | | | | | | neigh_GO_term =
nuclear_organization_and_biogenesis_: sce\overline{0}41\overline{1}1 (0.0)
```

```
| | | | | | | | | neigh GO term = peroxisome :
sce04111 (0.0)
| | | | | | | | | neigh_GO_term =
translation_regulator_activity_: sce04111 (0.0)
| | | | | | | | | neigh_GO_term =
electron_transport_: sce04111 (0.0)
| | | | | | | | neigh_GO_term =
extracellular_region_: sce04111 (0.0)
| | | | | | | | | neigh_GO_term =
isomerase_activity_: sce04111 (0.0)
| | | | | | | | | neigh_GO_term =
helicase_activity_: sce04111 (0.0)
| | | | | | | | | neigh_GO_term =
microtubule_organizing_center_: sce04111 (0.0)
| | | | | | | | | neigh_GO_term = cell_wall_:
sce04111 (0.0)
| | | | | | | | | neigh_GO_term =
cellular_homeostasis_: sce04111 (0.0)
| | \ | | \ | | | neigh_GO_term = lyase_activity_:
sce04111 (0.0)
| | | | | | | | | neigh_GO_term =
cellular_respiration_: sce04111 (0.0)
| | \ | | \ | | | neigh_GO_term = cell_budding_:
sce04111 (0.0)
| | | | | | | | | neigh_GO_term = motor_activity_:
sce04111 (0.0)
| | | | | | | start > 224304
| | | | | | | start <= 227393: sce00300
(28.83/13.65)
| | | | | | | | start > 227393: sce03010 (13.78)
| | | | | | start > 229906: sce04111 (47.49)
| | | | | start > 271901
| | | | | | strand = W: sce03030 (57.91/30.37)
| | | | | strand = C
| | | | | | | start > 276872
| | | | | | | | neigh_GO_term = mitochondrion_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term =
signal transducer activity : sce00562-(0.0)
| | | | | | | neigh_GO_term = biological_process_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term = cellular_component_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term =
oxidoreductase_activity_: sce00562 (0.0)
| | | | | | | | neigh_GO_term =
carbohydrate_metabolic_process_: sce0056\overline{2}}\mathrm{ (6.8/2.5)
| | | | | | | | neigh_GO_term = plasma_membrane_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term =
transporter_activity_: sce00562 (0.0)
| | | | | | | | neigh_GO_term = transport_: sce00632
(2.49/0.88)
```

```
| | | | | | | neigh_GO_term = molecular_function_:
sce00562 (1.51/0.77)
| | | | | | | | neigh_GO_term =
vitamin_metabolic_process_: sce00562 (0.0)
| | | | | | | | neigh_GO_term = cytoplasm_: sce00632
(6.88/3.58)
| | | | | | | | neigh_GO_term =
enzyme_regulator_activity_: sce00562 (0.0)
| | | | | | | | neigh_GO_term =
cell_wall_organization_and_biogenesis_: sce00562 (0.0)
| | | | | | | | neigh_GO_term =
signal transduction : sce00562 (0.0)
| | | | | | | | neigh_GO_term = membrane_: sce00562
(0.0)
| | | | | | | | neigh_GO_term = protein_binding_:
sce00562 (0.0)
| | | | | | | neigh_GO_term = sporulation_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term = hydrolase_activity_:
sce00632 (3.58/1.91)
| | | | | | | neigh_GO_term =
transferase_activity_: sce00562 (0.0)
| | | | | \ | | neigh GO term = response to stress :
sce00562 (3.68/1.67)
| | | | | | | | neigh_GO_term = nucleus_: sce00562
(5.61/1.09)
| | | | | | | neigh_GO_term =
protein_modification_process_: sce00562 (0.0)
| | | | | | | \ neigh_GO_term = vesicle-
mediated_transport_: sce00632 (5.05/1.68)
| | | | | | | | neigh_GO_term =
endomembrane_system_: sce00562 (0.0)
| | | | | | | | neigh GO term =
endoplasmic_reticulum_ sce04070 (1.3\overline{9/0.}
| | | | | | | neigh_GO_term =
phosphoprotein_phosphatase_activity_: sce00562 (0.0)
| | | | | | | | neigh_GO_term = translation_:
sce00562 (0.0)
| | | | | | | | neigh GO term = cell cortex :
sce00562 (0.0)
| | | | | | | neigh_GO_term = cytoskeleton_:
sce00562 (0.0)
| | | | | | | neigh_GO_term =
structural_molecule_activity_: sce005\overline{62 (0.0)}
| | | | | - | | \ neigh_GO_term = cytokinesis_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term =
DNA_metabolic_process_: sce00632 (1.65/0.95)
| | | | | | | neigh_GO_term = cell_cycle_:
sce00562 (0.0)
| | | | | | | neigh_GO_term = RNA_binding_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term =
anatomical_structure_morphogenesis_: sce00562 (0.0)
```

```
| | | | | | | neigh_GO_term = conjugation_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term = chromosome_:
sce00562 (2.11/0.7)
| | | | | | | | neigh_GO_term = DNA_binding_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term =
membrane_organization_and_biogenesis_: sce00562 (0.0)
| | | | | | | | neigh_GO_term = peptidase_activity_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term =
protein_kinase_activity_: sce00562 (0.0)
| | | | | | | | neigh_GO_term = nucleolus_: sce00562
(0.0)
| | | | | | | | neigh_GO_term =
ribosome_biogenesis_and_assembly_: sce00562 (0.0)
| | | | | - - | | ñeigh_GO_term = vacuole_: sce00562
(0.0)
| | | | | | | | neigh_GO_term =
RNA_metabolic_process_: sce00562 (0.0)
| | | | | | | | neigh_GO_term =
organelle_organization_and_biogenesis_: sce00562 (0.0)
| | | | | | | | neigh_GO_term = transcription_:
sce00562 (0.0)
| | | | | | | neigh_GO_term = Golgi_apparatus_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term =
cytoplasmic_membrane-bound_vesicle_: \overline{sce\overline{00562 (0.0)}}\mathbf{0}=0
| | | - | | | - | ne\overline{igh_GO_term = ribosome_: sce00562}
(0.0)
| | | | | | | | neigh_GO_term =
amino_acid_and_derivative_metabolic_process_: sce00562 (0.0)
| | | | | | | | neigh_GO_term =
generation_of_precursor_metabolites_and_ēnergy_: sce00562 (0.0)
| | | | | | | | neigh_GO_term =
nucleotidyltransferase_activity_
| | | | | | | | distance <= 4002: sce00632
(3.82/2.17)
| | | | | | | | distance > 4002: sce04070
(8.75/1.37)
| | | | | | | | neigh_GO_term =
transcription_regulator_activity_: sce00562 (0.0)
| | | | | | | | neigh_GO_term = ligase_activity_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term =
pseudohyphal_growth_: sce00562 (0.0)
| | | | | | | | neigh_GO_term = meiosis_: sce00562
(0.0)
| | | | | | | | neigh_GO_term = membrane_fraction_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term =
protein_catabolic_process_: sce00562 (0.00)
| | | | | | | | neigh_GO_term = cellular_bud_:
sce00562 (0.0)
```

```
| | | | | | | | neigh GO term =
cytoskeleton_organization_and_biogene\overline{sis__: sce00562 (0.0)}
| | | | | | | | neigh_GO_term =
mitochondrial_envelope_: sce00562 (0.0)
| | | | | | | neigh_GO_term =
site_of_polarized_growth_: sce00562 (\overline{0.0)}
| ` - | | 
lipid_metabolic_process_: sce00562 (0.0)
| | | | | | | | neigh_GO_term =
nuclear_organization_and_biogenesis_: sce00562 (0.0)
| | | | | | | | neigh_GO_term = peroxisome_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term =
translation_regulator_activity_: sce0\overline{0}56\overline{2}}\mathrm{ (0.0)
| | | | | | | | neigh_GO_term = electron_transport_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term =
extracellular region : sce00562 (0.0)
| | | | | \ | | neigh_GO_term = isomerase_activity_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term = helicase_activity_:
sce00562 (0.0)
| | | | | | | neigh GO term =
microtubule_organizing_center_: sce00562-(0.0)
| | | | | | | | neigh_GO_term = cell_wall_: sce00562
(0.0)
| | | | | | | neigh_GO_term =
cellular_homeostasis_: sce00562 (0.0)
| | \ | | \ | | neigh_GO_term = lyase_activity_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term =
cellular_respiration_: sce00562 (0.0)
| | | | | | | | neigh_GO_term = cell_budding_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term = motor_activity_:
sce00562 (0.0)
| | | start > 281848
| | | | start <= 340798
strand = W
| start <= 292781: sce03060 (14.45)
    | start > 292781: sce00361 (46.6/27.01)
    strand = C: sce03050 (20.22)
        start > 301413
        start <= 312471
        | start <= 303211: sce00190 (19.27)
        | start > 303211
        | | start <= 310122: sce03010 (41.12)
        | | start > 310122: sce00010 (12.69)
        start > 312471
        | start <= 316388
    | | neigh_GO_term = mitochondrion_:
sce00720 (0.0)
```

```
| | | | | | | | neigh_GO_term =
signal_transducer_activity_: sce00720-(0.0)
| | | | | | | | neigh_GO_term = biological_process_:
sce00720 (0.0)
| | | | | | | | neigh_GO_term = cellular_component_:
sce00720 (0.0)
| | | | | | | | neigh_GO_term =
oxidoreductase_activity_: sce00720 (0.0)
| | | | | | | | neigh_GO_term =
carbohydrate_metabolic_process_: sce00720 (0.0)
| | | | | | | | neigh_GO_term = plasma_membrane_:
sce00720 (0.0)
| | | | | | | | neigh_GO_term =
transporter_activity_: sce00720 (0.0)
| | | | | | | | neigh_GO_term = transport_: sce00720
(0.0)
| | | | | | | neigh_GO_term = molecular_function_:
sce00720 (0.0)
| | | | | | | | neigh GO term =
vitamin_metabolic_process_: sce00720 (0.0)
| | | | | | | | neigh_GO_term = cytoplasm_: sce00720
(0.0)
| | | | | | | | neigh_GO_term =
enzyme_regulator_activity
| | | | | | | | | neigh_num <= 1: sce00630
(3.16/1.97)
| | | | | | | | | neigh_num > 1: sce00720
(8.31/3.76)
| | | | | | | | neigh_GO_term =
cell_wall_organization_and_biogenesis_: sce00720 (0.0)
| | | | | | | | neigh_GO_term =
signal_transduction_: sce00720 (0.0)
| | | | | | | | neigh_GO_term = membrane_: sce00020
(11.0/7.42)
| | | | | | | neigh_GO_term = protein_binding_:
sce00720 (0.0)
| | | | | | | | neigh_GO_term = sporulation_:
sce00720 (0.0)
| | | | | | | neigh_GO_term = hydrolase_activity_:
sce00720 (0.0)
| | | | | | | | neigh_GO_term =
transferase_activity_: sce00720 (0.0)
| | | | | | | | neigh_GO_term = response_to_stress_:
sce00720 (0.0)
| | | | | | | | neigh_GO_term = nucleus_: sce00720
(0.0)
| | | | | | | | neigh_GO_term =
protein_modification_process_: sce00720 (0.0)
| | | | | \ | \ neigh_GO_term = vesicle-
mediated_transport_: sce00720 (0.0)
| | \ | | | | | neigh_GO_term =
endomembrane_system_: sce00720 (0.0)
| | | | | | | | neigh_GO_term =
endoplasmic_reticulum_: sce00720 (0.0)
```

```
| | | | | | | | neigh GO term =
phosphoprotein_phosphatase_activity_- sc\overline{e}00720 (0.0)
| | | | | | | | neigh_GO_term = translation_:
sce00720 (5.93/4.48)
| | | | | | | | neigh_GO_term = cell_cortex_:
sce00720 (0.0)
| | | | | | | neigh_GO_term = cytoskeleton_:
sce00720 (0.0)
| | | | | | | | neigh_GO_term =
structural_molecule_activity_: sce00620 (9.13/5.13)
| | | | | | | neigh_GO_term = cytokinesis_:
sce00720 (0.0)
| | | | | | | | neigh_GO_term =
DNA_metabolic_process_: sce00720 (0.0)
| | | | | | | | neigh_GO_term = cell_cycle_:
sce00720 (0.0)
| | | | | | | neigh_GO_term = RNA_binding_:
sce00720 (0.0)
| | | | | | | | neigh_GO_term =
anatomical_structure_morphogenesis_: sce\overline{0}0720 (0.0)
| | | | | | | | neigh_GO_term = conjugation_:
sce00720 (0.0)
| | | | | | | | neigh GO term = chromosome :
sce00720 (0.0)
| | | | | | | | neigh_GO_term = DNA_binding_:
sce00720 (0.0)
| | | | | | | | neigh_GO_term =
membrane organization and biogenesis : sce00720 (0.0)
| | \ | | | | | | neigh_GO_term = peptidase_activity_:
sce00720 (0.0)
| | | | | | | | neigh_GO_term =
protein_kinase_activity_: sce00720 (0.0)
| | | | | | | | neigh_GO_term = nucleolus_: sce00720
(0.0)
| | | | | | | | neigh_GO_term =
ribosome_biogenesis_and_assembly_: scé00720 (0.0)
| | | | | | | | neigh_GO_term = vacuole_: sce00630
(9.6/5.04)
| | | | | | | neigh GO term =
RNA_metabolic_process_: sce00720 (0.0)
| | | | | | | | neigh_GO_term =
organelle_organization_and_biogenesis_: sce00710 (7.51/5.18)
| | | | | | | | neigh_GO_term = transcription_:
sce00720 (0.0)
| | | | | | | | neigh_GO_term = Golgi_apparatus_:
sce00720 (0.0)
| | | | | | | | neigh_GO_term =
cytoplasmic_membrane-bound_vesicle_: sce00720 (0.0)
| | | | | | | neigh_GO_term = ribosome_: sce00720
(6.75/4.74)
| | | | | | | neigh_GO_term =
amino_acid_and_derivative_metabolic_process_: sce00720 (0.0)
| | | | | | | | neigh_GO_term =
generation_of_precursor_metabolites_and_energy_: sce00720 (0.0)
```

```
| | | | | | | | neigh GO term =
nucleotidyltransferase_activity_: sce\overline{0}07\overline{2}0 (0.0)
| | | | | | | | neigh_GO_term =
transcription_regulator_activity_: sce00720 (0.0)
| | | | | | | | neigh_GO_term = ligase_activity_:
sce00720 (0.0)
| | | | | | | | neigh_GO_term =
pseudohyphal_growth_: sce00720 (0.0)
| | | | | | | | neigh_GO_term = meiosis_: sce00720
(0.0)
| | | | | | | | neigh_GO_term = membrane_fraction_:
sce00720 (0.0)
| | | | | | | | neigh_GO_term =
protein_catabolic_process_: sce00720 (0.0)
| | | | | | | | neigh_GO_term = cellular_bud_:
sce00720 (0.0)
| | | | | | | | neigh GO term =
cytoskeleton organization and biogenesis}\mp@subsup{}{-}{-}\mathrm{ : sce00720 (0.0)
| | | | | | | | neigh GO_term =
mitochondrial_envelope_: sce00720 (0.\overline{0})
| | | | | | | neigh_GO_term =
site_of_polarized_growth_: sce00720 (0.0)
| \ | | | | ` | neigh GO term =
lipid_metabolic_process_: sce00720 (0.0)
| | | | | | | | neigh_GO_term =
nuclear_organization_and_biogenesis_: sce00720 (0.0)
| | | | | \ \ | neigh_GO_term = peroxisome_:
sce00720 (0.0)
| | | | | | | neigh_GO term =
translation_regulator_activity_: sce0\overline{0}72\overline{0}}\mathrm{ (0.0)
| | | | | | | neigh_GO_term = electron_transport_:
sce00720 (0.0)
| | | | | | | | neigh_GO_term =
extracellular_region_: sce00720 (0.0)
| | | | | \ | | neigh_GO_term = isomerase_activity_:
sce00720 (0.0)
| | | | | | | | neigh_GO_term = helicase_activity_:
sce00720 (0.0)
| | | | | | | | neigh_GO_term =
microtubule_organizing_center_: sce00720-(0.0)
| | | | | | | | neigh_GO_term = cell_wall_: sce00720
(0.0)
| | | | | | | | neigh_GO_term =
cellular_homeostasis_: sce00720 (0.0)
| | \ | | \ | | neigh_GO_term = lyase_activity_:
sce00720 (0.0)
| | | | | | | | neigh_GO_term =
cellular_respiration_: sce00720 (0.0)
| | | | | | | | neigh_GO_term = cell_budding_:
sce00720 (0.0)
| | | | | | | neigh_GO_term = motor_activity_:
sce00720 (0.0)
| | | | | | | start > 316388
    | strand = W
```





```
| | | | | | | | neigh_GO_term = vesicle-
mediated_transport_: sce04070 (0.0)
| | | | | | | | neigh_GO_term =
endomembrane_system_: sce04070 (0.0)
| | | | | | | | neigh_GO_term =
endoplasmic_reticulum_: sce04070 (0.0)
| | | - | |- | | neigh_GO_term =
phosphoprotein_phosphatase_activity_: sce04070 (0.0)
| | | | | | | | neigh_GO_term = translation_:
sce04070 (0.0)
| | | | | | | | neigh_GO_term = cell_cortex_:
sce04070 (0.0)
| | | | | | | neigh_GO_term = cytoskeleton_:
sce04070 (0.0)
| | | | | | | | neigh_GO_term =
structural_molecule_activity_: sce040\overline{70 (}0.0)
| | | | | | | | neigh_GO_term = cytokinesis_:
sce04070 (0.0)
| | | | | | | | neigh_GO_term =
DNA_metabolic_process_: sce00632 (5.49/2.08)
| | | | | | | | neigh_GO_term = cell_cycle_:
sce04070 (0.0)
| | | | | | | | neigh_GO_term = RNA_binding_:
sce04070 (0.0)
| | | | | | | | neigh_GO_term =
anatomical_structure_morphogenesis_: sce04070 (0.0)
| | | | | \ | | ne\overline{igh_GO_term = conjugation_:}
sce04070 (0.0)
| | | | | | | | neigh_GO_term = chromosome_:
sce04070 (3.42/1.77)
| | | | | | | | neigh_GO_term = DNA_binding_:
sce04070 (0.0)
| | | | | | | | neigh_GO_term =
membrane_organization_and_biogenesis_\overline{: s}\overline{c}e04070 (0.0)
| | \ | | | | | neigh_GO_term = peptidase_activity_:
sce04070 (0.0)
| | | | | | | | neigh_GO_term =
protein_kinase_activity_: sce04070 (0.0)
| | - | - | | | | neigh_GO_term = nucleolus_: sce04070
(0.0)
| | | | | | | | neigh_GO_term =
ribosome_biogenesis_and_assembly_: sce006}32\mathrm{ (2.46/1.2)
| | | | | | | | neigh_GO_term = vacuole_: sce04070
(0.0)
| | | | | | | | neigh_GO_term =
RNA_metabolic_process_: sce04070 (0.0)
| | | | | | | | neigh_GO_term =
organelle_organization_and_biogenesis_: sce04070 (0.0)
| | | | | | | | neigh_GO_term = transcription_:
sce04070 (0.73/0.42)
| | | | | | | | neigh_GO_term = Golgi_apparatus_:
sce04070 (0.0)
| | | | | | | | neigh_GO_term =
cytoplasmic_membrane-bound_vesicle_: sce04070 (0.0)
```

```
| | | | | | | | neigh GO term = ribosome : sce04070
(0.0)
| | | | | | | | neigh_GO_term =
amino_acid_and_derivative_metabolic_process_: sce04070 (0.0)
| | | | | | | | neigh_GO_term =
generation_of_precursor_metabolites_and_energy_: sce04070 (0.0)
| | | | | | | - | | neig}h G\overline{O}\mathrm{ term =
nucleotidyltransferase_activity_: sce\overline{0}05\overline{6}2 (8.04/3.45)
| | | | | | | | neigh_GO_term =
transcription_regulator_activity_: sceè00}\overline{6}32 (2.03/1.06
| | | | | | | | neigh_GO_term = ligase_activity_:
sce04070 (0.0)
| | | | | | | | neigh_GO_term =
pseudohyphal_growth_: sce04070 (0.0)
| | | | | | | | neigh_GO_term = meiosis_: sce04070
(0.0)
| | | | | | | | neigh_GO_term = membrane_fraction_:
sce04070 (0.0)
| | | | | | | | neigh_GO_term =
protein_catabolic_process_: sce04070 (0.00)
| | | | | | | | neigh_GO_term = cellular_bud_:
sce04070 (0.0)
| | | | | | | | neigh_GO_term =
cytoskeleton_organization_and_biogenesis__: sce04070 (0.0)
| | | \ | | | | - neigh_GO_term =
mitochondrial_envelope_: sce04070 (0.0)
| | | | | | | | neigh_GO_term =
site_of_polarized_growth_: sce04070 (\overline{0.0)}
| \ | | | | \ | neigh_GO_term =
lipid_metabolic_process_: sce04070 (10}.3\overline{7}/4.85
| | | | | | | | neigh_GO_term =
nuclear_organization_and_biogenesis_: sce04070 (0.0)
| | | | | \ \ | neig}h_GO_term = peroxisome_:
sce04070 (0.0)
| | | | | | | | neigh_GO_term =
translation_regulator_activity_: sce0\overline{4}07\overline{0}}\mathrm{ (0.0)
| | | | | | | | neigh_GO_term = electron_transport_:
sce04070 (0.0)
| | | | | | | | neigh_GO_term =
extracellular_region_: sce04070 (0.0)
| | | | | | | | neigh_GO_term = isomerase_activity_:
sce04070 (0.0)
| | | | | | | | neigh_GO_term = helicase_activity_:
sce04070 (0.0)
| | | | | | | | neigh_GO_term =
microtubule_organizing_center_: sce04070-(0.0)
| | | | | | | | neigh_GO_term = cell_wall_: sce04070
(0.0)
| | | | | | | | neigh_GO_term =
cellular_homeostasis_: sce04070 (0.0)
| | \ | | \ | | neigh_GO_term = lyase_activity_:
sce04070 (0.0)
| | | | | | | | neigh_GO_term =
cellular_respiration_: sce04070 (0.0)
```



```
        strand = W
        start <= 946803
        | neigh_num <= 1: sce00643 (42.28/29.72)
        | neigh_num > 1: sce00220 (46.41/34.87)
        start > 946803: sce01030 (40.15/15.22)
        strand = C
        start <= 946803: sce00740 (18.22)
        start > 946803: sce00030 (23.9)
        958335
        start <= 1009006
        strand = W
        start <= 1004000: sce00380 (40.16/19.83)
        start > 1004000: sce00620 (12.27)
        strand = C
        | start <= 969676
        | neigh_num <= 1: sce00380 (8.86/2.23)
        | neigh_num > 1: sce00680 (22.45/8.27)
        start > 9\overline{6}9676: sce00500 (13.45)
        start > 1009006
        start <= 1056547
        start <= 1045007: sce03060 (16.16)
        start > 1045007: sce00600 (39.57)
        start > 1056547
        start <= 1058810: sce00190 (16.77)
        start > 1058810: sce00220 (21.95)
    start > 1062787
        strand = W
            start <= 1085062
            start <= 1067727: sce00563 (37.18/17.19)
            start > 1067727: sce03022 (21.43)
        start > 1085062
            start <= 1117121
        | start <= 1108699: sce00460 (37.87/20.33)
        start > 1108699: sce04111 (13.98)
        start > 1117121: sce01031 (38.22/17.25)
        strand = C
        start <= 1129583: sce04120 (28.21/13.83)
        start > 1129583
        | distance <= 2089: sce00970 (12.43/1.62)
        | distance > 2089: sce00330 (10.7/2.7)
    start > 1153620
        start <= 1369782
        start <= 1234210
        | start <= 1213896
            start <= 1184740
            | start <= 1183292: sce00240 (8.76)
            | start > 1183292: sce00400 (12.12)
            start > 1184740: sce00650 (69.86/53.06)
        start > 1213896
            start <= 1228603: sce00190 (19.5)
            start > 1228603: sce00010 (13.78)
        start > 1234210
            strand = W
        start <= 1322197
```




## Decision Tree Generated for Chromosome Five:

```
start <= 257957
    start <= 160549
        start <= 107260
        strand = W
        | start <= 43252: sce00530 (27.47)
        start > 43252
        | start <= 86937
            start <= 68792: sce00190 (18.42)
            start > 68792: sce04111 (5.95)
            start > 86937: sce00190 (36.02)
        strand = C
        start <= 68792
            start <= 51539: sce03030 (14.38)
            start > 51539
            | start <= 58378: sce03010 (16.57)
            start > 58378: sce00260 (11.67)
        start > 68792: sce00513 (38.72/17.31)
            start > 107260
            start <= 133120
                start <= 128825
                        start <= 116167: sce00240 (17.71)
            start > 116167: sce04140 (20.63)
        start > 128825
            start <= 131772: sce04120 (19.59)
            start > 131772: sce00500 (21.87)
            start > 133120
        start <= 153519: sce01030 (45.68/15.15)
        start > 153519
            start <= 159117: sce00051 (18.61)
        start > 159117: sce00230 (21.48/9.77)
```



|  |  |
| :---: | :---: |
| । | neigh_GO_term = mitochondrion_: sce00220 |
| (10.91/8.25) |  |
| 1111 | neigh_GO_term = cytoplasmic_membrane-bound_vesicle_: |
| sce00340 (0.0) |  |
| (12.18/10.5) |  |
|  |  |
| \| | | | | neigh GO term = motor activity : sce00340 (0.0) |
| \| | | | | neigh_GO_term $=$ cell_cycle_: sce00340 (0.0) |
| \| | | | | neigh_GO_term = vacuole : sce00340 (0.0) |
| \| | | | | neigh_GO_term = peptidase_activity_: sce00340 (0.0) |
| 1 \| | | | neigh_GO_term = sporulation_: sce00340 (0.0) |
| । \| | | | neigh_GO_term = mitochondriāl_envelope_: sce00340 |
| (0.0) |  |
| \| | | | | neigh_GO_term = DNA_metabolic_process_: sce00340 |
| (0.0) |  |
| \| | | | | neigh GO_term = isomerase activity : sce00340 (0.0) |
| \| | | | | neigh_GO_term = carbohydrate_metabolic_process_: |
| sce00340 (0.0) |  |
| \| | | | | neigh_GO_term |
| cell wall organization and biogenesis : sce00340 (0.0) |  |
| । † । ${ }^{-}$ | neigh_̄̄O_tèrm = proteīn_binding_: sce00340 (0.0) |
| 1 \| | | | neigh_GO_term = nucleolus_: sce000340 (0.0) |
| \| | | | | neigh_GO_term = nucleotidyltransferase_activity_: |
| sce00340 (0.0) _ot |  |
| \| | | | | neigh_GO_term = transcription_: sce00340 (0.0) |
| 1 \| | | | neigh_GO_term = RNA_binding_: sce00340 (0.0) |
| \| | | | | neigh_GO_term = RNA_metabolic_process_: sce00340 |
| (0.0) |  |
| 1 \| | | neigh_GO_term = ribosome_biogenesis_and_assembly_: |
| sce00340 (0.0) |  |
| \| | | | | neigh GO term = ribosome : sce00340 (0.0) |
|  | neigh GO-term = structural molecule activity : |
| sce00120 (20.39/16.33) |  |
| \| | | | | neigh GO term = translation : sce00380 (36.35/29.25) |
| \| | | | | neigh_GO_term $=$ response_to_stress_: sce00340 (0.0) |
| \| | | | | neigh_GO_term $=$ membrane_- ${ }^{-}$sce0034 $\overline{0}$ (0.0) |
| 1 \| | | | neigh_GO_term = hydrolase__activity_: sce00340 (0.0) |
| \| | | | | neigh_GO_term |
| organelle_organization and biogenesis_: sce00220 (11.91/9.81) |  |
| \| | | | | neigh_Ḡo_term = lyase_activity_: sce00340 (0.0) |
| \| | | | | neigh_GO_term = |
| amino_acid_and_derivativè_metabolic_process_: sce00340 (0.0) |  |
| \| | | | | neigh_GO_term = electron_transport_: sce00340 (0.0) |
| \| | | | | neigh_GO_term = Golgi_apparatus_: sce00340 (0.0) |
| \| | | | | neigh_GO_term = vitamin metabolic_process : sce00340 |
| (0.0) - - - - |  |
| \| | | | | neigh_GO_term = membrane_fraction_: sce00340 (0.0) |
| \| | | | | neigh_GO_term = DNA_binding_: scē00340 (0.0) |
| \| | | | | neigh_GO_term = protein_catabolic_process_: sce00340 |
| (0.0) |  |
| 1 I | neigh_GO_term = vesicle-mediated_transport_: |
| sce00340 (0.0) |  |






```
| | | | | neigh GO term = signal transducer activity :
sce00010 (0.0)
| | | | start > 546812: sce00670 (13.73)
```


## Decision Tree Generated for Chromosome Six:

```
start <= 103121
| start <= 54695
| | start <= 15431
| | | start <= 10969: sce00750 (14.32)
| | | start > 10969
| | | | neigh_GO_term = cellular_component_: sce00650
(10.24/7.42)
| | | | neigh_GO_term = molecular_function_: sce00650
(15.99/7.29)
| | | | neigh_GO_term = biological_process_: sce00260
(7.47/5.07)
| | | | neigh_GO_term = vacuole_: sce00120 (0.0)
| | | | neigh_GO_term = vitamin_metabolic_process_: sce00260
(5.26/3.87)
| | | | neigh_GO_term = protein_binding_: sce00120 (0.0)
| | | | neigh_GO_term = oxidore\overline{ductase_āctivity_: sce00051}
(7.2/2.68)
| | | | 
| | | | neigh_GO_term = transporter_activity_: sce00120
(15.7/7.74)
\begin{tabular}{|c|c|c|}
\hline 1 | | & & neigh_GO_term \(=\) transport_: sce00310 (6.95/4.68) \\
\hline 1 | & I & neigh_GO_term \(=\) membrane_: sce00120 (6.27/4.32) \\
\hline 1 | & 1 | & neigh_GO_term = transferase_activity_: sce00120 \\
\hline \multicolumn{3}{|l|}{(0.0) - - - -} \\
\hline 1 | & I & neigh_GO_term = response_to_stress_: sce00120 (0.0) \\
\hline 1 | & 1 | & neigh_GO_term = DNA_binding_: sce00120 (0.0) \\
\hline 1 | & 1 | & neigh_GO_term = nucleus_: sçe00120 (0.0) \\
\hline 1 | | & & neigh_GO_term = transcription_regulator_activity_: \\
\hline \multicolumn{3}{|l|}{sce00120 (0.0)} \\
\hline 1 | & + & neigh_GO_term = transcription_: sce00120 (0.0) \\
\hline 1 | & 1 | & neigh_GO_term = Golgi_apparatus_: sce00120 (0.0) \\
\hline 1 | | & 1 & neigh_GO_term = vesicle-mediated_transport_: \\
\hline \multicolumn{3}{|l|}{sce00120 (0.0) - - -} \\
\hline 1 | & I & neigh_GO_term = cytoplasm_: sce00120 (0.0) \\
\hline 1 | & 1 | & neigh_GO_term = enzyme_regulator_activity_: sce00120 \\
\hline \multicolumn{3}{|l|}{(0.0) - - - - -} \\
\hline 1 | & 1 | & neigh_GO_term = signal_transduction_: sce00120 (0.0) \\
\hline 1 | & I & neigh_GO_term = isomerase_activity_: sce00120 (0.0) \\
\hline 1 | & 1 I & neigh_GO_term = mitochondrion_: sce00120 (0.0) \\
\hline 1 | & | & neigh_GO_term = peptidase_activity_: sce00120 (0.0) \\
\hline 1 | & 1 | & neigh_GO_term = cell_cortex_: sce00120 (0.0) \\
\hline 1 | & 1 | & neigh_GO_term = cellular_bud_: sce00120 (0.0) \\
\hline 1 | & 1 | & neigh_GO_term = cytoskeleton_: sce00120 (0.0) \\
\hline | & 1 | & neigh_GO_term = site_of_polarized_growth_: sce00120 \\
\hline
\end{tabular}
```

|  |  |
| :---: | :---: |
|  |  |
| , | neigh_GO_term = |
| cell_wall_organization_and_biogenesis_: sce00120 (0.0) |  |
| \| | | | | neigh_GO_term = cytokinesis_: sce00120 (0.0) |
| , | neigh_GO_term = microtubule_organizing_center |
| sce00120 (0.0) |  |
| \| | | | | neigh_GO_term = cell_cycle_: sce00120 (0.0) |
| 1 \| | | | neigh_GO_term = meiosis_: sce00120 (0.0) |
| 1 \| | | neigh_GO_term = hydrolase_activity_: sce00120 (0.0) |
| I | neigh_GO_term = protein_modification_process_: |
| sce00120 (0.0) |  |
| \| | | neigh_GO_term = nucleotidyltransferase_activity_: |
| sce00120 (0.0) |  |
| \| | | | | neigh_GO_term = |
| organelle_organization_and_biogenesis_: sce00120 (0.0) |  |
| \| | | | | neigh_Ḡ_term = ribosome_: sce00120 (0.0) |
| 1 \| | | neigh_GO_term = translation_: sce00120 (0.0) |
| 1 \| | | neigh_GO_term = protein_kinase_activity_: sce00120 |
| (0.0) |  |
| 1 \| | | neigh_GO_term = lipid_metabolic_process_: sce00120 |
| (0.0) |  |
| 1 \| | | | neigh_GO_term |
| amino_acid_and_derivative_metabolic_process_: sce00120 (0.0) |  |
| \| | | | | neigh_GO_term = membrane_fraction_: sce00120 (0.0) |
| 1 \| | | | neigh_GO_term = signal_transducer_activity_: |
| sce00120 (0.0) |  |
| \| | | | | neigh_GO_term = conjugation_: sce00120 (0.0) |
| 1 \| | | neigh_GO_term = endoplasmic_reticulum_ sce00120 |
| (0.0) |  |
| 1 \| | | neigh_GO_term = DNA_metabolic_process_: sce00120 |
| (0.0) |  |
| 1 \| | | neigh_GO_term = ligase_activity_: sce00120 (0.0) |
| 1 \| | | neigh_GO_term = RNA_binding_: sce00120 (0.0) |
| 1 \| | | neigh_GO_term = RNA_metabolic_process_: sce00120 |
| (0.0) |  |
| 1 \| | | neigh_GO_term = carbohydrate_metabolic_process_ |
| sce00120 (0.0) |  |
| \| | | | | neigh_GO_term = chromosome_: sce00120 (0.0) |
| 1 \| | | neigh_GO_term = protein_catabolic_process_: sce00120 |
| (0.0) |  |
| 1 \| | | neigh_GO_term = |
| membrane_organization_and_biogenesis_: sce00120 (0.0) |  |
| - | neigh_GO_term = nucleolus_: sce00120 (0.0) |
| 1 \| | | | neigh_GO_term = helicase_activity_: sce00120 (0.0) |
| 1 \| | | neigh_GO_term = ribosome_biogenesis_and_assembly_: |
| sce00120 (0.0) |  |
|  | neigh_GO_term = |
| nuclear_organization_ānd_biogenesis_: sce00120 (0.0) |  |
| 1 \| | neigh_GO_term = |
| generation_of_precursor_metabolites_and_energy_ sce00120 (0.0) |  |
| \| | neigh_GO_term = |
| cytoskeleton_o | ganization_and_biogenesis_: sce00120 (0.0) |

```
| | | | neigh GO term =
phosphoprotein_phosphātase__activity_: sce00120 (0.0)
| | | | neigh_GO_term = cellular_respiration_: sce00120
(0.0)
| | | neigh_GO_term = electron_transport_: sce00120 (0.0)
(0.0)
| | | | neigh_GO_term = lyase_activity_: sce00120 (0.0)
| | start > 15431
    | strand = W: sce00561 (12.99)
    strand = C
        | start <= 44392: sce00051 (16.95)
        start > 44392
                        start <= 45560: sce00632 (26.49/11.67)
                        start > 45560: sce01430 (9.67)
    start > 54695
        start <= 95008
        strand = W: sce04010 (17.02)
        strand = C
        | start <= 87232
        | start <= 82578
        | start <= 64931: sce03010 (10.69)
        | start > 64931: sce04111 (15.52)
        start > 82578: sce00563 (13.36)
        start > 87232
        neigh_strand = W: sce00400 (8.48/1.17)
        neigh_strand = C: sce00970 (18.53/6.68)
    start > 95008
        neigh_num <= 1: sce00020 (31.87/21.53)
        neigh_num > 1: sce00260 (37.49/25.01)
    ne
    start <= 226950
    strand = W
        start <= 153388
        | start <= 119424
        | start <= 116139
            | neigh_num <= 1: sce04120 (9.88/3.56)
            | neigh_num > 1: sce04111 (25.26/6.04)
            start > 1\overline{1}6139: sce04111 (17.72)
        start > 119424: sce03050 (17.76)
        start > 153388
        | start <= 220094
            start <= 184490: sce04070 (19.38)
            start > 184490
            | neigh_strand = W: sce00920 (7.05/1.54)
            neigh_strand = C: sce00450 (11.73/3.89)
        strand = C
        start <= 204738
        | start <= 119424: sce00251 (31.85/12.5)
        start > 119424
        start <= 184490: sce00500 (29.08)
        start > 184490: sce00340 (15.06)
        start > 204738
```

```
| | | | start <= 220094: sce04111 (38.96)
| start > 220094
    | start <= 224757
    | start <= 223425: sce03010 (15.18)
    | start > 223425: sce00190 (12.83)
    start > 224757: sce04111 (19.36)
start > 226950
    start <= 255036
    start <= 252492
        start <= 245153: sce00760 (10.14)
        start > 245153: sce03050 (28.0)
        start > 252492
            distance <= 339: sce00521 (6.95/1.41)
            distance > 339
        | neigh_strand = W: sce00521 (36.85/26.99)
        | neigh_strand = C: sce00051 (15.33/7.82)
    start > 255036
        neigh strand = W: sce00920 (29.37/19.54)
        neigh_strand = C
        neigh_num <= 1: sce00910 (7.81/4.85)
        neigh_num > 1: sce00450 (9.52/3.91)
```


## Decision Tree Generated for Chromosome Seven:

```
start <= 216101
| start <= 23935
| | start <= 14910
| | | start <= 14157
| | | | neigh_GO_term = cellular_component_: sce01031
(8.77/2.93)
| | | | neigh_GO_term = molecular_function_: sce00561
(5.7/3.39)
| | | | neigh_GO_term = biological_process_: sce00561
(3.33/1.81)
| | | | neigh_GO_term = transferase_activity_: sce01031
(0.0)
| | | | neigh_GO_term = protein_modification_process_:
sce01031 (0.0)
| | | | neigh_GO_term = peptidase_activity_: sce01031 (0.0)
| | | neigh_GO_term = mitochondrion_: scē00051 (2.29/1.27)
| | | | neigh_GO_term = oxidoreductase__activity_: sce01031
(6.22/3.11)
| | | | neigh_GO_term =
generation_of_precursor_metabolites_and_energy_: sce00561
(5.19/2.08)
| | | | neigh_GO_term = cytoplasm_: sce00051 (6.5/1.36)
| | | | neigh_GO_term = transcription_regulator_activity_:
sce00561 (1.18/0.48)
| | | | neigh_GO_term = signal_transduction_: sce01031
(4.06/1.67)
| | | | neigh_GO_term = plasma_membrane_: sce01031 (0.0)
| | | | neigh_GO_term = transpōrter_activity_: sce01031
(0.0)
```









```
| | | | | | | | | neigh_GO_term =
molecular_function_: sce00680 (10.29/4.18)
| | | | | | | | | neigh_GO_term =
biological_process_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
transferase activity : sce00670 (0.0)
| | | | | \ | | | neigh GO term =
protein_modification_process_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
peptidase_activity_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term = mitochondrion_:
sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
oxidoreductase_activity_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
generation_of_precursor_metabolites_and_energy_: sce00670 (0.0)
| | | | | | | | | - neig}h_GO_tērm = cytoplasm_:
sce00670 (1.75/0.79)
| | | | | | | | | neigh_GO_term =
transcription_regulator_activity_: sce006\overline{70 (0.0)}
| | | | | | | | | neigh_GO_term =
signal_transduction_: sce00670 (0.0)
| | | | | - | | | | neigh GO term =
plasma_membrane_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
transporter_activity_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term = transport_:
sce00670 (0.0)
| | | | | | | | neigh_GO_term = nucleus_:
sce00670 (1.4/0.67)
| | | | | | | | | neigh_GO_term = DNA_binding_:
sce00670 (0.0)
| | | | | | | | | neigh_GO_term = transcription_:
sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
anatomical_structure_morphogenesis_: sce0\overline{0}67\overline{0}}\mathrm{ (0.0)
| | | | | | | | | neigh_GO_term =
helicase_activity_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term = meiosis_:
sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
enzyme_regulator_activity_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
ribosome_biogenesis_and_assembly_: sce006\overline{70 (0.0)}
| | | | | - | | ` neigh GO term = chromosome :
sce00670 (0.0)
| | | | | | | | neigh_GO_term = cell_cycle_:
sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
hydrolase_activity_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
endomembrane_system_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
membrane_organization_and_biogenesis_: sc\overline{e}00\overline{6}70 (0.0)
```

```
| | | | | | | | | neigh GO term =
nuclear_organization_and_biogenesis_: sce\overline{0}06\overline{7}0 (0.0)
| | | | | | | | | neigh_GO_term =
ligase_activity_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term = translation_:
sce00670 (5.48/1.24)
| | | | | | | | | neigh_GO_term =
DNA_metabolic_process_: sce00670 (2.22/0.47)
| | | | | | | | | neigh_GO_term =
RNA_metabolic_process_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
protein_binding_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
endoplasmic_reticulum_: sce00670 (1.61/0.38)
| | | | | | | | | neigh_GO_term = cellular_bud_:
sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
site_of_polarized_growth_: sce00670 (0.0)
| | | | | | \ | | neigh_GO_term = cytokinesis_:
sce00670 (0.0)
| | | | | | | | | neigh_GO_term = RNA_binding_:
sce00670 (0.0)
| | | | | | | | | neigh GO term =
phosphoprotein_phosphatase_activity_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term = membrane_:
sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
protein_catabolic_process_: sce00670 (0.0)
| | - | | | | | | neigh_GO_term =
Golgi_apparatus_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term = vesicle-
mediated_transport_: sce00680 (3.61/1.46)
| | | | | | | | | neigh GO term =
organelle_organization_and_biogenesis_: s\overline{ce0}0\overline{0}70 (0.0)
| | | | | | | | | neigh_GO_term = motor_activity_:
sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
cytoskeleton_organization_and_biogenesis_: sce00670 (0.0)
| | | \ | | | | | | neigh_GO_term = vacuole_:
sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
cytoplasmic_membrane-bound_vesicle_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
structural_molecule_activity_: sce00680 (\overline{4.5}\overline{8}/0.72)
| | | | | - | | \ | neigh_GO_term =
protein_kinase_activity_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term = peroxisome_:
sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
lipid_metabolic_process_: sce00680 (6.57/\overline{1}.6\overline{2})
| | | | | | | | | neigh_GO_term =
amino_acid_and_derivative_metabolic_process_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term = sporulation_:
sce00670 (0.0)
```

```
| | | | | | | | neigh GO term = lyase activity :
sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
pseudohyphal_growth_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
cellular respiration : sce00670 (0.0)
| | | | \ | | | neigh_GO_term = ribosome_:
sce00670 (6.52/0.97)
| | | | | | | | | neigh_GO_term =
electron_transport_: sce00670 (0.0)
| | | | | | | | | neigh GO term =
cell_wall_organization_and_biogenesis_: s\overline{ce0}0}670\mathrm{ (0.0)
| | | | | | | _ | | nëigh_GO_term =
microtubule_organizing_center_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term = nucleolus_:
sce00670 (0.58/0.11)
| | | | | | | | | neigh_GO_term =
isomerase activity : sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
carbohydrate_metabolic_process_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
translation_regulator_activity_: sce00670 (0.0)
| | | - | | | | | | neigh GO term =
nucleotidyltransferase_activity_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
signal_transducer_activity_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
extracellular_region_: sce00670 (0.0)
| | | | | \ | | | neigh_GO_term = conjugation_:
sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
response_to_stress_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
vitamin_metabolic_process_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term = cell_budding_:
sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
membrane_fraction_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term = cell_wall_:
sce00670 (0.0)
| | | | | | | | | neigh_GO_term =
mitochondrial_envelope_: sce00670 (0.0)
| | | | | | | | | neigh_GO_term = cell_cortex_:
sce00670 (0.0)
| | | | | | | | | neigh_GO_term = cytoskeleton_:
sce00670 (0.0)
| | | start > 272526: sce03010 (10.86)
start > 277623
| start <= 299731: sce04111 (37.86/15.57)
start > 299731
| start <= 317345
| | start <= 310970: sce03010 (13.54)
| start > 310970: sce04130 (7.75)
    start > 317345: sce04111 (20.04)
```



```
start > 511549
| start <= 648911
| | start <= 527333
start <= 513163: sce00272 (46.47/27.15)
start > 513163
    | start <= 525233: sce00410 (63.03/39.68)
    | start > 525233: sce00190 (9.51)
    start > 527333
    strand = C
    | start <= 581440
    | start <= 558872
        | start <= 534462: sce03010 (20.61)
        start > 534462: sce00510 (23.1)
        start > 558872: sce00030 (23.54)
        start > 581440
            start <= 615969: sce00230 (13.53)
            | start > 615969: sce03010 (32.66)
        strand = W
        start <= 581440
        | start <= 548268: sce04010 (21.33/9.33)
        start > 548268
        | | start <= 558872: sce03010 (13.7)
        | | start > 558872: sce04010 (11.51)
        start > 581440: sce00100 (18.11)
        start > 648911
        start <= 718897
        start <= 654638
            start <= 652985: sce00010 (20.61)
            start > 652985: sce00380 (50.08/32.38)
        start > 654638
            start <= 672190
            start <= 668193: sce04111 (16.73)
            start > 668193: sce00290 (24.86/11.2)
        start > 672190: sce04111 (68.06)
        718897
        start <= 739949
            start <= 726978: sce03010 (11.46)
            start > 726978: sce00252 (29.85/11.61)
        start > 739949
            start <= 761397: sce03050 (28.64)
            start > 761397: sce03010 (15.52)
            start > 787784
        start <= 958213
            start <= 867525
                start <= 842556
            strand = C: sce00271 (50.08/29.01)
            strand = W
            start <= 798548: sce00271 (43.43/25.21)
            start > 798548: sce00260 (51.32/22.65)
            start > 842556
            start <= 856306
            | start <= 848428: sce00900 (30.01/11.83)
            | start > 848428: sce00240 (43.75/16.53)
            start > 856306
```



## Decision Tree Generated for Chromosome Eight:

```
start <= 323412
| start <= 126515
| | strand = C
| | | start <= 81612
    | start <= 39484
        start <= 39484
        | start <= 36023: sce03010 (11.24)
    start > 36023: sce00561 (10.81)
        start > 38506: sce04130 (11.56)
        start > 39484: sce00230 (31.4/11.76)
        start > 81612
        | start <= 114912
        start <= 107821
        | start <= 97933: sce04010 (16.1)
        | start > 97933: sce00600 (18.01)
        start > 107821: sce04010 (17.89)
        start > 114912: sce00100 (20.18)
        strand = W
        start <= 81612
        | start <= 75409: sce03010 (10.64)
        start > 75409
            neigh_strand = W: sce00520 (13.2/6.44)
            neigh_strand = C: sce00040 (25.22/11.13)
            start > 81612
            | start <= 114912
                start <= 104272: sce03010 (7.77)
        start > 104272: sce00190 (9.22)
        start > 114912: sce03010 (20.32)
    start > 126515
        | start <= 286772
| | start <= 159431
        start <= 143989
            start <= 127774: sce00970 (25.29/9.23)
            start > 127774
            | start <= 141395
            | | neigh_num <= 1: sce00252 (16.23/8.59)
            | neigh_num > 1
                | nēigh_strand = W: sce00330
(5.19/1.94)
| | | | | | | | neigh_strand = C: sce00220
(6.33/2.52)
| | | | | | start > 141395
| | | | | | | | | start <= 143551
(10.06/3.67)
| | | | | | | | neigh_num > 1: sce00252 (19.29/7.61)
    | | | start > 1\overline{4}3551: sce00330 (33.82/15.52)
    start > 143989
        | start <= 148662: sce03010 (10.35)
        | start > 148662: sce00260 (17.58)
        start > 159431
        start <= 209699
        start <= 170337
        | strand = C
```





```
| | | | | | start <= 451327: sce00010 (10.75)
| | | | start > 451327: sce00030 (17.65)
start > 470960
    start <= 499079
    strand = C
    | neigh_strand = W: sce01031 (16.59/4.28)
        neigh_strand = C: sce00563 (14.37/6.04)
        strand = \overline{W}
        start <= 484845: sce00900 (33.72/15.5)
        start > 484845: sce03050 (13.47)
        start > 499079
        strand = C
        | start <= 501143: sce00230 (13.16)
        | start > 501143: sce03010 (20.31)
        strand = W
        start <= 517532: sce00290 (80.01/41.52)
        start > 517532
            start <= 552099: sce00740 (17.35/8.32)
                        start > 552099: sce00230 (8.72)
```


## Decision Tree Generated for Chromosome Nine:

```
start <= 155219
| start <= 99416
| | start <= 41444
| | | start <= 37385: sce00500 (28.55)
| | | start > 37385
| | | | neigh_GO_term = cellular_component_: sce00071 (0.0)
neigh_GO_term = helicase_activity_: sce00071 (0.0)
neigh_GO_term = biological_process_: sce00071
neigh GO term = molecular function
| neigh_strand = W: sce00592 (6.19/3.93)
| neigh_strand = C: sce00120 (16.77/10.98)
neigh_GO_term = cytoplasm_: sce00280 (10.68/6.92)
neigh_GO-term = hydrolase_activity_: sce00071 (0.0)
neigh_GO_term = transport_: sce00071 (0.0)
neigh_GO_term = vesicle-mediated_transport_:
sce00071 (0.0)
| | | | neigh_GO_term = membrane_: sce00071 (0.0)
| | | | neigh_GO_term = transporĒer_activity_: sce00071
(0.0)
| | | | neigh_GO_term = extracellular_region_: sce00071
(0.0)
| | | neigh_GO_term = mitochondrion_: sce00592 (8.21/6.04)
| | | | neigh_GO_term = carbohydrate_metabolic_process_:
sce00071 (0.0)
| | | | neigh_GO_term = peroxisome_: sce00071 (0.0)
| | | | neigh_GO_term = transferase__activity_: sce00071
(0.0)
| | | | neigh_GO_term = lipid_metabolic_process_: sce00071
(0.0)
| | | | neigh_GO_term = cellular_bud_: sce00281 (12.97/7.26)
```




```
| | | start <= 187629
| strand = C
    | start <= 166412: sce00051 (23.4)
    start > 166412: sce00300 (13.38)
    strand = W
    start <= 166412: sce04111 (20.52)
    start > 166412: sce00500 (15.58)
    start > 187629
    strand = C
    start <= 220697: sce03050 (25.09)
    start > 220697: sce00260 (21.03)
    strand = W: sce00260 (23.15/9.01)
start > 222487
    start <= 312903
    | start <= 257061
        strand = C
        | start <= 232366: sce03010 (19.06)
        | start > 232366
        | start <= 240705: sce00240 (42.26/19.32)
        | | start > 240705: sce03010 (14.63)
        strand = W: sce00561 (8.56)
        start > 257061
            start <= 307927
            | strand = C
            | | start <= 274925: sce00530 (23.36)
            | start > 274925: sce04111 (8.65)
            strand =W: sce04111 (43.74/13.51)
            start > 307927: sce00230 (52.47/27.2)
    start > 312903
    | strand = C: sce00340 (17.3)
    strand = W
        start <= 316766: sce03010 (14.93)
        start > 316766: sce00051 (42.44/26.9)
        start > 326101
        start <= 415614
        start <= 348502
        start <= 339341: sce00071 (21.69)
        start > 339341: sce04130 (19.77)
        start > 348502
        start <= 393672
        | start <= 374303: sce00240 (46.78/26.61)
        start > 374303: sce00500 (10.43)
        start > 393672
            start <= 410804: sce00230 (27.96)
            start > 410804
                start <= 414676: sce00630 (27.31/7.26)
                start > 414676: sce00230 (19.87)
    start > 415614
        start <= 420733: sce00310 (32.7/13.41)
        start > 420733
        neigh_num <= 1
        neigh_strand = W: sce00590 (3.94/1.42)
        neigh_strand = C: sce00480 (10.19/2.23)
        neigh_num > 1: sce00590 (13.3/4.37)
```


## Decision Tree Generated for Chromosome Ten:

```
start <= 424103
| start <= 105008
| | start <= 26086
| | | strand = C: sce00500 (31.05)
| | | strand=W N | neigh_GO_term = cellular_component_: sce00564
(27.35/14.49)
| | | | neigh_GO_term = helicase_activity_: sce00624 (0.0)
| | | | neigh_GO_term = biologic\overline{al_process_}_
| | | | | nēigh_strand = W
| | | | | | nēigh_num <= 1: sce00624 (4.32/3.03)
| | | | | | neigh_num > 1: sce00632 (7.77/4.31)
| | | | | | neigh_strand = C | neigh_num <= 1: sce00564 (12.58/9.65)
| | | | | | neigh_num > 1: sce00903 (11.02/7.37)
| | | | neigh_GO_term= molecular_function_: sce00624
(15.94/10.93)
| | | | neigh_GO_term = hydrolase_activity_: sce00632
(17.88/12.95)
| | | neigh_GO_term = transport_: sce00624 (16.23/9.82)
| | | | neigh_GO_term = vesicle-mediated_transport_:
sce00624 (0.0)
| | | | neigh_GO_term = plasma_membrane_: sce00632
(7.91/5.3)
| | | | neigh_GO_term = transporter_activity_: sce00624
(3.82/2.91)
| | | | neigh_GO_term = transferase_activity_: sce00624
(0.0)
| | | | neigh_GO_term = cytoplasm_: sce00624 (4.37/3.41)
| | | | neigh_GO_term = endoplasmic_reticulum_: sce00624
(0.0)
```



```
organelle_organization_and_biogenesis_: sce00624 (0.0)
| | | neigh_GO_term = mitochondrion_: sce00624 (0.0)
| | | | neigh_GO_term = RNA_binding_:- sce00624 (0.0)
| | | | neigh_GO_term = cel\overline{lular_respiration_: sce00624}
(0.0)
| | | | neigh_GO_term = nucleus_: sce00624 (0.0)
| | | | neigh_GO_term = ribosome_biogenesis_and_assembly_:
sce00624 (0.0)
| | | | neigh_GO_term = RNA_metabolic_process_: sce00624
(0.0)
| | | | neigh_GO_term = Golgi_apparatus_: sce00624 (0.0)
| | | | neigh_GO_term = site_of_polarized_growth_: sce00624
(0.0)
| | | | neigh_GO_term =
cell_wall_organization_añd_biogenesis_: sce00624 (0.0)
```



```
| | | | neigh_GO_term = translation_regulator_activity_:
sce00624 (0.0)
| | | | neigh_GO_term = cytoplasmic_membrane-bound_vesicle_:
sce00624 (0.0)
| | | | neigh_GO_term =
phosphoprotein_phosphatase_activity_: sce00624 (0.0)
| | | | neigh_GO_tērm = signal_transducer_activity_:
sce00624 (0.0)
| | | | neigh_GO_term = cell_cortex_: sce00624 (0.0)
    neigh_GO_term = cytokinesis_: sce00624 (0.0)
    neigh_GO_term =
nuclear organization \overline{and biogenesis : sce00624 (0.0)}
| | | | neigh_GO_term = sporulation_: sce00624 (0.0)
| | | | neigh_GO_term =
membrane_organization_and_biogenesis_: sce00624 (0.0)
| | | | neigh_GO_term = cellular_bud_: sce00624 (0.0)
| | | | neigh_GO_term = protein_\overline{catabolic_process_: sce00624}
(0.0)
| | | | neigh_GO_term = microtubule_organizing_center_:
sce00624 (0.0)
| | | | neigh_GO_term = mitochondrial_envelope_: sce00624
(0.0)
| | | | neigh_GO_term = cell_budding_: sce00624 (0.0)
| | | | neigh_GO_term = pseud\overline{dohyphal_growth_: sce00624 (0.0)}
| | start > 26086
| | | start <= 79261
| | | | start <= 58813
| | | | | start <= 36919: sce04111 (13.7)
| | | | | start > 36919
| | | | | | neigh_GO_term = cellular_component_:
sce00720 (0.0)
| | | | | | neigh_GO_term = helicase_activity_: sce00720
(0.0)
| | | | | | neigh GO term = biological process :
sce00720 (0.0)
| | | | | | neigh_GO_term = molecular_function_:
sce00020 (13.57/4.75)
| | | | | | neigh_GO_term = hydrolase_activity_:
sce00720 (0.0)
| | | | | | neigh_GO_term = transport_: sce00630
(1.53/0.73)
| | | | | | neigh_GO_term = vesicle-mediated_transport_:
sce00720 (0.0)
| | | | | | neigh_GO_term = plasma_membrane_: sce00720
(0.0)
| | | | | | neigh_GO_term = transporter_activity_:
sce00720 (4.57/2.84)
| | | | | | neigh_GO_term = transferase_activity_:
sce00720 (0.0)
| | | | | | neigh_GO_term = cytoplasm_: sce00720
(8.91/3.59)
| | | | | | neigh_GO_term = endoplasmic_reticulum_:
sce00720 (0.0)
| | | | | | neigh_GO_term = ribosome_: sce00720 (0.0)
```









```
| | | start > 667131: sce00510 (65.1/41.06)
| | start > 667638
| | | start <= 705738
| | | | strand = C
    start <= 690518: sce00300 (23.56/9.38)
    start > 690518: sce03010 (5.88)
        strand = W
    neigh_num <= 1
    | neigh_strand = W: sce00280 (7.36/3.61)
    | neigh_strand = C: sce00290 (12.76/7.39)
    neigh_num }\mp@subsup{}{}{-}>1: sce00770 (29.76/13.05
    start > 70573\overline{8}
    start <= 722809: sce00500 (38.49/14.82)
    start > 722809: sce00051 (79.29/50.29)
```


## Decision Tree Generated for Chromosome Eleven:

```
start <= 327131
    start <= 194865
        start <= 100676
            strand = C
        start <= 75539
        start <= 39164: sce00400 (22.05)
        start > 39164: sce04130 (11.62)
        start > 75539
            start <= 78651: sce00260 (43.58/13.65)
            start > 78651: sce00190 (18.0)
        strand = W
        start <= 39164
        | start <= 25216: sce00240 (25.77)
        | start > 25216: sce04120 (12.94)
        start > 39164
            start <= 96762: sce00220 (18.33)
            start > 96762: sce00061 (12.1)
        start > 100676
        start <= 166549
            strand = C
            start <= 114632: sce04010 (21.35)
            start > 114632
            start <= 140696: sce01031 (26.51/13.06)
            start > 140696: sce00010 (16.52)
        strand = W
            start <= 107321: sce00230 (33.4/16.15)
            start > 107321
            start <= 158619: sce03010 (36.44)
            start > 158619: sce00530 (7.37)
        start > 166549
        start <= 179672
            start <= 174218
            | start <= 171134: sce00020 (38.56/13.03)
            start > 171134: sce03050 (6.46)
```



|  | neigh_GO_term = protein_kinase_activity_: sce00030 |
| :---: | :---: |
| (7.86/3.24) |  |
| 1 \| | | | neigh_GO_term = peroxisome_: sce00500 (0.0) |
| 1 \| | | neigh_GO_term = ligase_activity_: sce00500 (0.0) |
| 1 \| | | neigh_GO_term = enzyme_regulator_activity_: sce00500 |
| (0.0) |  |
| \| | | | neigh_GO_term = |
| phosphoprotein_phosphātase_activity_: sce00500 (0.0) |  |
| + | neigh_GO_term = |
| cell_wall_organization_and_biogenesis_: sce00500 (0.0) |  |
| \| | | | | neigh_GO_term = cellular_bud_: sce00500 (0.0) |
| 1 \| | | | neigh_GO_term = transcription_: sce00500 (0.0) |
| 1 \| | | | neigh_GO_term = transcription_regulator_activity_: |
| sce00500 (0.0) |  |
| $1 \quad 1 \quad 1$ | neigh_GO_term = pseudohyphal_growth_: sce00500 (0.0) |
| 1 \| | | neigh_GO_term = endomembrane_system_: sce00500 (0.0) |
| 1 \| | | neigh_GO_term = lipid_metabolic_process_: sce00500 |
| (0.0) |  |
| 1 \| | | neigh_GO_term = transferase_activity_: sce00500 |
| (0.0) |  |
| (0.0) | neigh_GO_term = structural_molecule_activity_: |
| sce00500 (0.0) |  |
| 1 \| | | | neigh_GO_term = signal_transducer_activity_: |
| sce00500 (0.0) |  |
| 1 \| | | | neigh_GO_term = conjugation_: sce00500 (0.0) |
| 1 \| | | neigh_GO_term = vacuole_: sce00500 (0.0) |
| 1 \| | | neigh_GO_term = cellular_homeostasis_: sce00500 |
| (0.0) |  |
| 1 \| | | neigh_GO_term = cytoplasmic_membrane-bound_vesicle_: |
| sce00500 (0.0) |  |
| 1 I l \| | neigh_GO_term = Golgi_apparatus_: sce00500 (0.0) |
| 1 \| | | neigh_GO_term = cell_wall_: sce00500 (0.0) |
| 1 \| | | neigh_GO_term = endoplasmic_reticulum_: sce00500 |
| (0.0) |  |
| 1 \| | | neigh_GO_term = peptidase_activity_: sce00500 (0.0) |
| 1 \| | | neigh_GO_term = isomerase_activity_: sce00500 (0.0) |
| \| | | | | neigh_GO_term = carbohydrate_metabolic_process_: |
| sce00500 (0.0) |  |
| 1 I l \| | neigh_GO_term = electron_transport_: sce00500 (0.0) |
| 1 \| | | | neigh_GO_term = nucleotidyltransferase_activity_: |
| sce00500 (0.0) - - - |  |
| 1 I l \| | neigh_GO_term = chromosome_: sce00500 (0.0) |
| 1 \| | | neigh_GO_term = cytoskeleton_: sce00500 (0.0) |
| 1 \| | | neigh_GO_term = cell_cycle_: sce00500 (0.0) |
| 1 \| | | neigh_GO_term = cell_cortex_: sce00500 (0.0) |
| 1 \| | | neigh_GO_term = motor_activity_: sce00500 (0.0) |
| 1 \| | | | neigh_GO_term = |
| organelle_organization_and_biogenesis_: sce00500 (0.0) |  |
| \| | | | | neigh_GO_term = DNA_binding_: sce00500 (0.0) |
| 1 \| | | neigh_GO_term = cell_budding_: sce00500 (0.0) |
|  | neigh_GO_term = cellular_respiration_: sce00500 |
| (0.0) |  |
| 1 \| | | neigh_GO_term = mitochondrial_envelope_: sce00500 |
| (0.0) |  |




| $\begin{aligned} & \text { \| \| \| \| } \\ & (0.0) \end{aligned}$ |  |  |
| :---: | :---: | :---: |
|  |  |  |
|  |  | neigh_GO_term = carbohydrate_metabolic_process_: |
| sce00710 (0.0) |  |  |
| \| | | \| | neigh_GO_term = electron_transport_: sce00710 |
| (0.0) |  |  |
| 1 \| | | \| | neigh_GO_term = |
| nucleotidyltransferase_activity_: sce00710 (0.0) |  |  |
| , | \| | neigh_GO_term = chromosome_: sce00710 (0.0) |
| 1 \| | \| | neigh_GO_term = cytoskeleton_: sce00710 (0.0) |
| 1 \| | \| | neigh_GO_term = cell_cycle_: sce00401 (5.3/4.09) |
| 1 \| | \| | neigh_GO_term = cell_cortex_: sce00710 (0.0) |
| 1 \| | I | neigh_GO_term = motor_activity_: sce00710 (0.0) |
| 1 \| | | \| | | neigh_GO_term = |
| organelle_organization_and_biogenesis_: sce00710 (0.0) |  |  |
| $\|1\|$ | - \| | neīgh_Ḡ_term = DNAA_binding_: sce00710 (0.0) |
| 1 \| | \| | neigh_GO_term = cell_budding_: sce00710 (0.0) |
| 1 \| | \| | neigh_GO_term = cellular_respiration_: sce00710 |
| (0.0) - - - - |  |  |
|  |  | neigh_GO_term = mitochondrial_envelope_: |
| sce00710 | (0.0) |  |
|  |  | neigh_GO_term = translation_regulator_activity_: |
| sce00710 | (0.0) |  |
| \| | \| | neigh_GO_term = helicase_activity_: sce00710 |
| (0.0) |  |  |
| 1 \| | | \| | | neigh_GO_term |
| nuclear_organization_and_biogenesis_: sce00710 (0.0) |  |  |
| \| | | | \| | | neigh_GO_term = cytokinesis_: sce00710 (0.0) |
| \| |  | neigh_GO_term = microtubule_organizing_center_: |
| sce00710 | (0.0) |  |
| \| | \| | neigh_GO_term = |
| generation_of_precursor_metabolites_and_energy_: sce00710 (0.0) |  |  |
| \| | | | \| | | neigh_GO_term = |
| membrane_organization_and_biogenesis_: sce00710 (0.0) |  |  |
| 1 \| | - \| | nēigh_GO_term = extracellular_region_: sce00710 |
| (0.0) |  |  |
| 1 \| | \| | neigh_GO_term = meiosis_: sce00710 (0.0) |
| 1 \| | | \| | t > 237180: sce00530 (35.17/17.16) |
| t > 245017 |  |  |
| \| | | start <= 278767 |  |  |
| 1 \| | \| | t $<=248564: \operatorname{sce} 04111$ (9.95) |
| 1 \| | \| | start > 248564 |
| 1 \| | \| | neigh_num <= 1 |
| 1 \| | \| | \| distance <= 2075: sce00020 (22.51/13.03) |
| 1 \| | \| | \| distance > 2075: sce00620 (4.94/1.23) |
| 1 \| | I | neigh_num > 1: sce00720 (60.61/46.96) |
| 1 \| |  | 278767 |
| 1 \| | \| | nd $=\mathrm{C}: ~ \mathrm{sce} 00010$ (62.18/42.08) |
| 1 \| | \| | nd $=\mathrm{W}$ |
| 1 \| | \| | start $<=284674$ : sce00190 (15.35) |
| 1 \| | I | start > 284674 |
| 1 \| | \| | start $<=314456$ : sce00240 (21.0/8.01) |
| - | \| | | start > 314456: sce04010 (8.7) |

```
strand = C
    start <= 425518
        | start <= 389526
        | start <= 353137: sce04070 (46.96/28.58)
        start > 353137
        | start <= 385401: sce00620 (13.88)
        | start > 385401: sce00590 (27.43/9.76)
        start > 389526
        start <= 392783: sce00240 (14.95)
        start > 392783
        | start <= 396891: sce04111 (34.29/15.85)
        start > 396891
                start <= 407632: sce00190 (19.85)
                start > 407632: sce04120 (13.57)
    start > 425518
        start <= 456697
        start <= 439028
        start <= 431549
        | start <= 428194: sce00600 (19.28)
        | start > 428194: sce04130 (14.24)
        start > 431549: sce00450 (36.12/19.5)
        start > 439028
        | neigh_strand = W: sce00632 (98.89/83.78)
        neigh_strand = C: sce00624 (69.53/54.99)
        start > 456697
        start <= 506037: sce00565 (38.33/17.96)
        start > 506037
        | start <= 571254: sce00600 (16.97)
        start > 571254: sce03010 (16.37)
    strand = W
        start <= 559308
        start <= 353137
        | start <= 329810: sce03022 (23.87)
        start > 329810: sce03030 (48.11/28.26)
        start > 353137
        start <= 369534
        neigh_strand = W
        | distance <= 1705: sce00040 (9.47/4.05)
        | distance > 1705: sce00500 (17.22/6.31)
        neigh strand = C: sce00040 (13.39/6.88)
        start > 3\overline{6}9534
        start <= 551299
            | start <= 407632: sce03022 (15.59)
            | start > 407632: sce03010 (34.69)
            start > 551299: sce03022 (16.33)
        start > 559308
        start <= 619447
        start <= 571254: sce00860 (19.95)
        start > 571254
            start <= 590037: sce00670 (17.59)
            start > 590037: sce04010 (17.03)
        start > 619447: sce00620 (61.31/37.4)
```


## Decision Tree Generated for Chromosome Twelve:

```
start <= 487290
| start <= 286560
| | strand = C
| | | start <= 198086
| | | | start <= 166536
| | | | | start <= 80204
| | | | | | start <= 48628: sce03010 (18.25)
| | start > 48628: sce04140 (16.62)
| start > 52589
    | start <= 53930: sce00020 (36.43/13.81)
    | start > 53930: sce00563 (21.82/6.09)
start > 80204
    start <= 111574
    | neigh_strand = W: sce00252 (8.96/2.09)
    | neigh_strand = C: sce00970 (24.33/9.88)
    start > 1\overline{1}1574
    | start <= 141072: sce00190 (17.16)
    | start > 141072: sce04010 (29.47/9.66)
    166536
        start <= 196473: sce04130 (22.21)
        start > 196473
        | neigh_GO_term = cellular_component_:
sce00350 (0.0)
| | | | | | neigh_GO_term = helicase_activity_: sce00350
(0.0)
| | | | | | neigh_GO_term = biological_process_:
sce00350 (0.0)
| | | | | | neigh_GO_term = molecular_function_:
sce00350 (0.0)
| | | | | | neigh_GO_term = plasma_membrane_: sce00350
(0.0)
| | | | | | neigh_GO_term = transporter_activity_:
sce00350 (0.0)
| | | | | | neigh_GO_term = transport_: sce00350 (0.0)
| | | | | | neigh_GO_term = transferase_activity_:
sce00350 (0.0)
| | | | | neigh_GO_term = cytoplasm_: sce00350
(15.89/10.98)
| | | | | | neigh_GO_term =
amino_acid_and_derivative_metabolic_process_: sce00350 (0.0)
| | | | | | neigh_GO_term = mitochondrion_: sce00350
(0.0)
| | | | | neigh_GO_term = oxidoreductase_activity_:
sce00350 (0.0)
| | | | | | neigh_GO_term = endoplasmic_reticulum_:
sce00350 (0.0)
| | | | | | neigh_GO_term =
transcription_regulator_activity_: sce00350 (0.0)
| | | | | - neigh_GO_term = membrane_: sce00350 (0.0)
| | | | | | neigh_GO_term = vacuole_` sce00350 (0.0)
```



```
| | | | | | neigh GO term =
phosphoprotein_phosphatase_ac\overline{t}iv\overline{i}ty_: sce00350 (0.0)
| | | | | | neigh_GO_term =
cell_wall_organization_and_biogenesis_: sce00350 (0.0)
| | | | neigh_GO_term = chromosome_: sce00350 (0.0)
| | | | neigh_GO_term = DNA_binding_: sce00350 (0.0)
microtubule_organizing_center_: sce00350 (0.0)
| | | | | | neigh_GO_term = cell_cycle_: sce00350 (0.0)
| | | | | | neigh_GO_term = enzyme_regulator_activity_:
sce00350 (0.0)
| | | | | | neigh_GO_term = signal_transducer_activity_:
sce00350 (0.0)
| | | | | | neigh_GO_term = signal_transduction_:
sce00350 (0.0)
| | | | | | neigh_GO_term = transcription_: sce00350
(0.0)
| | | | | | neigh_GO_term = Golgi_apparatus_: sce00272
(19.73/15.34)
| | | | | | neigh_GO_term = peroxisome_: sce00350 (0.0)
| | | | | neigh_GO_term = meiosis_: Sce00350 (0.0)
| | | | neigh_GO_term = cell_wall_: sce00350 (0.0)
| neigh GO term = lyase activity : sce00350
(0.0)
| | | | | | neigh_GO_term =
generation_of_precursor_metabolites_and_energy_: sce00350 (0.0)
| | | | | | neigh_GO_term = cellular_homeostasis_:
sce00350 (0.0)
| | | | | | neigh_GO_term = ligase_activity_: sce00350
(0.0)
| | | | | | neigh_GO_term = peptidase_activity_:
sce00350 (0.0)
| | | | | | neigh_GO_term =
translation_regulator_activit\overline{y}}:
| | | | | | neigh_GO_term =
carbohydrate_metabolic_process_: sce00350 (0.0)
| | | | | | neigh_GO_term = endomembrane_system_:
sce00350 (0.0)
| | | | | | neigh_GO_term = cytokinesis_: sce00350 (0.0)
| | | | | | neigh_GO_term =
membrane_organization_and_biogenesis_: sce00350 (0.0)
| | | | | | neigh_GO_term = isomerase_activity_:
sce00350 (0.0)
| | | | | | neigh_GO_term = membrane_fraction_: sce00350
(0.0)
| | | | | | neigh_GO_term =
nuclear_organization_and_biogenesis_: sce00350 (0.0)
| | | | | | neigh_GO_term = vitamin_metabolic_process_:
sce00350 (0.0)
| | | | | | neigh_GO_term = conjugation_: sce00350 (0.0)
| | | | | neigh_GO_term = extracellular_region_:
sce00350 (0.0)
| | | | | | neigh_GO_term =
nucleotidyltransferase_activity_: sce00350 (0.0)
```








```
| | | | | neigh_GO_term = mitochondrial_envelope_:
sce00460 (0.0)
| | | | | | neigh_GO_term = cellular_respiration_:
sce00460 (0.0)
| | | | | | neigh_GO_term =
organelle_organization_and_biogenesis_: sce00460 (0.0)
| | | | | | -
sce00460 (0.0)
| | | | | | neigh_GO_term =
phosphoprotein_phosphatase_activity_: sce00460 (0.0)
| | | | neigh_GO_term =
cell_wall_organization_and_biōge\overline{nesis_: sce00430 (9.9/7.41)}
| | | | | - ne\overline{igh_GO_term = chromosome_: sce00460 (0.0)}
| | | | | | neigh_GO_term = DNA_binding__: sce00460 (0.0)
| | | | | | neigh_GO_term =
microtubule_organizing_center_: sce00460 (0.0)
| | | | | | neigh_GO_term = cell_cycle_: sce00460 (0.0)
| | | | | | neigh_GO_term = enzymēe_regūlator_activity_:
sce00460 (0.0)
| | | | | | neigh_GO_term = signal_transducer_activity_:
sce00460 (0.0)
| | | | | | neigh_GO_term = signal_transduction_:
sce00460 (0.0)
| | | | | | neigh_GO_term = transcription_: sce00460
(0.0)
| | | | | | neigh_GO_term = Golgi_apparatus_: sce00460
(0.0)
| | | | | | neigh_GO_term = peroxisome_: sce00460 (0.0)
| | | | | neigh_GO_term = meiosis_: sce00460 (0.0)
| | | | neigh_GO_term = cell_wall_: sce00450
(4.82/3.34)
| | | | | | neigh_GO_term = lyase_activity_: sce00460
(0.0)
| | | | | | neigh GO term =
generation_of_precursor_metabolites_and_energy_: sce00460 (0.0)
| | | | | | neigh_GO_term = cellular_homeostasis_:
sce00460 (0.0)
| | | | | | neigh_GO_term = ligase_activity_: sce00460
(0.0)
| | | | | | neigh_GO_term = peptidase_activity_:
sce00460 (0.0)
| | | | | | neigh_GO_term =
translation_regulator_activity_: sce00460 (0.0)
| | | | | | neigh_GO_term =
carbohydrate metabolic process :- sce00460 (0.0)
| | | | | | neigh_\overline{GO_term = endomembrane_system_:}
sce00460 (0.0)
| | | | | neigh_GO_term = cytokinesis_: sce00460 (0.0)
| | | | | neigh_GO_term =
membrane organization and bio\overline{genésis : sce00460 (0.0)}
| | | | | | neigh_GO_term = isomerase_activity_:
sce00460 (0.0)
| | | | | | neigh_GO_term = membrane_fraction_: sce00460
(0.0)
```




## Decision Tree Generated for Chromosome Thirteen:

```
start <= 396378
| start <= 104162
| | strand = C
| | | start <= 20536
| | | | neigh_GO_term = mitochondrion_: sce00280 (6.42/3.2)
    neigh_GO_term = helicase_activity_: sce00650 (0.0)
    neigh_GO_term = biological_process_: sce00650 (0.0)
    neigh_GO_term = membrane_fraction_: sce00650 (0.0)
    neigh_GO_term = protein_\overline{b}inding_:- sce00650 (0.0)
    neigh_GO_term = cellular_homeostasis_: sce00650
(0.0)
| | | | neigh_GO_term = endoplasmic_reticulum_: sce00280
(7.97/3.61)
| | | | neigh_GO_term = oxidoreductase_activity_: sce00650
(0.0)
| | | | neigh_GO_term = protein_modification_process_:
sce00650 (0.0)
| | | | neigh_GO_term = cytoplasm_: sce00280 (2.87/1.27)
| | | | neigh_GO_term = molecular_function_
| | | | | neig\overline{h_num <= 1: sce000}72 (2.89\overline{/1.15)}
| | | | | neigh_num > 1: sce00650 (20.12/6.03)
| | | | neigh_GO_term = cellular_component_: sce00280
(7.21/2.55)
| | | | neigh GO_term = nucleus
    | dístānce <= 2869: s\overline{ce00650 (5.9/2.69)}
    | distance > 2869: sce00280 (5.45/2.3)
    neigh_GO_term = response_to_stress_: sce00650
(3.18/1.01)
| | | | neigh_GO_term = mitochondrial_envelope_: sce00650
```




```
| | | | start <= 45063
    neigh_strand = W: sce00632 (26.96/14.8)
    neigh_strand = C: sce04070 (17.39/7.87)
    start > 45063
    start <= 58687: sce00240 (23.35)
    start > 58687: sce00500 (25.27)
    start > 70624
    start <= 86739: sce03022 (14.27)
    start > 86739: sce00920 (66.68/46.78)
    start > 104162
    start <= 302484
    start <= 228937
        | start <= 145876
            start <= 133475
            | strand = C
            | start <= 118898: sce00100 (14.82)
                    | start > 118898: sce03010 (8.31)
                    strand = W: sce00561 (16.58)
            start > 133475: sce04111 (30.16)
        start > 145876
            start <= 146482: sce03010 (17.32)
            start > 146482
            | start <= 208860
            | strand = C
            | | start <= 164176: sce00230 (16.13)
                    start > 164176
                    start <= 167308: sce00620
                    (13.58)
```










```
| | | | | | neigh GO term =
ribosome_biogenesis_and_assem\overline{bly_}: sce00030 (0.0)
| | | | | | neigh_GO_term =
amino_acid_and_derivative_metabolic_process_: sce00030 (0.0)
| | | | | | neigh_GO_term = enzyme_regulator_activity_:
sce00030 (0.0)
| | | | | | neigh_GO_term = peptidase_activity_:
sce00030 (0.0)
| | | | | | neigh_GO_term = protein_catabolic_process_:
sce00030 (0.0)
| | | | | | neigh_GO_term = translation_: sce00030 (0.0)
| | | | | | neigh_GO_term = lyase_activíty_: sce00030
(0.0)
| | | | | | neigh_GO_term =
generation_of_precursor_metabolites_and_energy_: sce00030 (0.0)
| | | | | | neigh_GO_term = isomerase_activity_:
sce00030 (0.0)
| | | | | | neigh_GO_term = vesicle-mediated_transport_:
sce00030 (0.0)
| | | | | | neigh_GO_term = endomembrane_system_:
sce00030 (0.0)
| | | | | | neigh_GO_term = cellular_bud_: sce00030
(0.0)
| | | | | | neigh_GO_term = cytoplasmic_membrane-
bound_vesicle_: sce00030 (0.0)
| | | | | | neigh_GO_term =
cell_wall_organization_and_biogenesis_: sce00030 (0.0)
| \ | | | | - ne\overline{igh_GO_term = sporulation_: sce00030 (0.0)}
| | | | | | neigh_GO_term =
phosphoprotein_phosphatase_activity_: sce00030 (0.0)
| | | | | neigh_GO_term = conjugation_: sce00030 (0.0)
| | | | | neigh_GO_term = cell_cortex_: sce00030 (0.0)
carbohydrate metabolic_process_: sce00030 (0.0)
| | | | | | neigh_GO_term = peroxisome_: sce00010
(2.73/1.0)
| | | | | | neigh_GO_term = vitamin_metabolic_process_:
sce00030 (0.0)
| | | | | | neigh_GO_term =
cytoskeleton organization and biōgenesis : sce00030 (0.0)
| | | | | | neigh_GO_term = vacuole_: sce00030 (0.0)
| | | | | | neigh_GO_term = signal_transduction_:
sce00030 (0.0)
| | | | | neigh_GO_term = extracellular_region_:
sce00030 (0.0)
| | | | | | neigh_GO_term = cell_wall_: sce00030 (0.0)
| | | | | | neigh_GO_term = pseudohyphal_growth_:
sce00030 (0.0)
| | | | | | neigh_GO_term = site_of_polarized_growth_:
sce00030 (0.0)
| | | | | | neigh_GO_term = cytokinesis_: sce00030 (0.0)
| | | | | | neigh_GO_term =
membrane_organization_and_biogenesis_: sce00030 (0.0)
```


$\mid$
$\mid$
$\mid$
$\mid$
$\mid$$|$



| $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | start $<=901708: \operatorname{sce} 03050(21.14)$ |  |
| start $>901708: \operatorname{sce} 00010(15.62)$ |  |  |  |

## Decision Tree Generated for Chromosome Fourteen:

```
start <= 529943
| strand = W
| | start <= 155101
| | | start <= 140879
| | | | start <= 87897
| | | | | start <= 66517: sce03030 (12.22)
        | | start > 66517: sce00240 (25.56)
        | start > 82806: sce04111 (20.6)
        start > 87897
            start <= 117349: sce00271 (34.33/14.59)
            start > 117349: sce04070 (25.27/11.33)
        start > 140879
        | start <= 148213: sce03030 (66.72/42.21)
        start > 148213: sce04111 (27.09)
        start > 155101
        start <= 331324
        | start <= 237664
            start <= 197944
            | start <= 164624: sce00790 (16.41)
            | start > 164624: sce00272 (41.19/18.07)
            start > 197944: sce00230 (50.06/21.96)
        start > 237664
            start <= 276503: sce00530 (13.12)
            start > 276503
            | start <= 302682: sce03010 (13.41)
            | start > 302682
            | | start <= 317673: sce04120 (19.93/3.44)
                            start > 317673: sce03010 (11.59)
            start > 331324
        | start <= 444317
            start <= 406360
            | start <= 380833
            | | start <= 352416: sce04010 (9.79)
            start > 352416: sce00230 (8.91)
            start > 380833
            | neigh_num <= 1: sce00620 (18.26/6.25)
            | neigh_num > 1: sce00630 (28.96/13.06)
            start > 40636\overline{0}
            | start <= 413641
            | start <= 412773
                | neigh_GO_term = nucleus_: sce03020
(0.77/0.35)
| | | | | | | | neigh_GO_term = helicase_activity_:
sce00230 (3.08/1.87)
| | | | | | | | neigh_GO_term =
DNA_metabolic_process_: sce03020 (0.0)
```

```
| | | | | | | neigh GO term = vacuole : sce00230
(3.46/1.15)
| | | | | | | neigh_GO_term = molecular_function_:
sce03020 (6.33/2.82)
| | | | | | | | neigh_GO_term = biological_process_:
|ce00240 (7.7/3.02) | | | | | | neigh GO term = cellular_component :
sce03020 (0.0)
| | | | | | | | neigh_GO_term =
vitamin_metabolic_process_: sce03020 (0.0)
| | | | | | | | neigh_GO_term = protein_binding_:
sce03020 (0.0)
| | | | | | | | neigh_GO_term =
oxidoreductase_activity_: sce03020 (0.0)
| | | | | | | | neigh_GO_term = cellular_bud_:
sce03020 (0.0)
| | | | | | | | neigh_GO_term = hydrolase_activity_:
sce03020 (0.0)
| | | | | | | neigh_GO_term = cytokinesis_:
sce03020 (0.0)
| | | | | | | | neigh_GO_term =
transcription_regulator_activity_: scé03020 (0.0)
| | | | | - | | neigh GO term =
protein_modification_process_: sce030\overline{20}
| | | | | | | neigh_GO_term = membrane_fraction_:
sce03020 (0.0)
| | | | | | | | neigh_GO_term = peroxisome_:
sce03020 (0.0)
| | | | | | | neigh_GO_term =
signal_transduction_: sce03020 (0.0)
| | | | | | | | neigh_GO_term =
mitochondrial_envelope_: sce03020 (0.0)
| | | | | | | | neigh GO term =
enzyme_regulator activity_ sce03020 (0.\overline{0)}
| | | | | | | | neigh_GO_term =
transporter_activity_: sce03020 (0.0)
| | | | | | | | neigh_GO_term =
organelle_organization_and_biogenesis_: sce03020 (0.0)
| | | | | | - | | neigh_GO_term = transport_: sce03020
(0.0)
| | | | | | | | neigh_GO_term =
transferase_activity_: sce03020 (0.0)
| | | | | | | | neigh_GO_term =
membrane_organization_and_biogenesis_: sce03020 (0.0)
| | \ | | | | | | neigh_GO_term = membrane_: sce03020
(0.0)
| | | | | | | neigh_GO_term = conjugation_:
sce03020 (0.0)
| | | | | | | | neigh_GO_term = cytoplasm_: sce03020
(8.39/2.28)
| | | | | | | | neigh_GO_term =
endoplasmic_reticulum_: sce03020 (0.0)
| | | | | | | | neigh_GO_term = plasma_membrane_:
sce03020 (0.0)
```

```
| | | | | | | | neigh_GO_term = cell_wall_: sce03020
(0.0)
| | | | | | | | neigh_GO_term =
structural_molecule_activity_: sce03020 (0.0)
| | | | | | | | neigh_GO_term =
cell_wall_organization_and_biogenesis_: sce03020 (0.0)
| \ | | | | - | | | neigh_GO_term = mitochondrion_:
sce00240 (4.84/1.66)
| | | | | | | | neigh_GO_term = RNA_binding_:
sce03020 (0.0)
| | | | | | | | neigh_GO_term =
RNA_metabolic_process_: sce03020 (0.0)
| | | | | | | | neigh_GO_term = lyase_activity_:
sce03020 (0.0)
| | | | | | | | neigh_GO_term =
amino_acid_and_derivative_metabolic_process_: sce03020 (0.0)
| |}\mp@subsup{|}{}{-}\mp@subsup{|}{}{-}| | | | nei\overline{gh_GO_term = transcription_:
sce03020 (0.0)
| | | | | | | neigh_GO_term = chromosome_:
sce03020 (0.0)
| | | | | | | | neigh_GO_term = DNA_binding_:
sce03020 (0.0)
| | | | | | | | neigh_GO_term = ribosome_: sce03020
(0.0)
| | | | | | | | neigh_GO_term =
protein_catabolic_process_: sce03020 (0.0)
| | | | | | | | neigh_GO_term = cell_cycle_:
sce03020 (0.0)
| | | | | | | | neigh_GO_term = response_to_stress_:
sce03020 (0.0)
| | | | | | | | neigh_GO_term = nucleolus_: sce03020
(0.0)
| | | | | | | | neigh_GO_term =
ribosome_biogenesis_and_assembly_: sc\overline{e}00\overline{2}30 (5.88/3.01)
| | | | | | | neigh_GO_term = translation_:
sce03020 (0.0)
| | | | | | | | neigh_GO_term =
protein_kinase_activity_: sce03020 (0.0)
| | - | - | | - | | neigh_GO_term = meiosis_: sce03020
(0.0)
| | | | | | | | neigh_GO_term =
site_of_polarized_growth_: sce03020 (0.0)
| | | | | | \ | neigh_GO_term =
nucleotidyltransferase_activity_: sce\overline{0}30\overline{2}0 (0.0)
| | | | | | - | | neigh_GO_term = cell_cortex_:
sce03020 (0.0)
| | | | | | | | neigh_GO_term = cytoskeleton_:
sce03020 (0.0)
| | | | | | | neigh_GO_term = sporulation_:
sce03020 (0.0)
| | | | | | | | neigh_GO_term = vesicle-
mediated_transport_: sce03020 (0.0)
| | | | | | | | neigh_GO_term = isomerase_activity_:
sce03020 (0.0)
```

```
| | | | | | | | neigh_GO_term = Golgi_apparatus_:
sce03020 (0.0)
| | | | | | | | neigh_GO_term =
signal_transducer_activity_: sce03020 (0.0)
| | | | | | | | neigh_GO_term =
lipid_metabolic_process_: sce03020 (0.0)
| |- | | | | - | | neigh GO term =
cytoplasmic_membrane-bound_vesicle_: sce03020 (0.0)
| | | | | | | | neigh_GO_term =
pseudohyphal_growth_: sce03020 (0.0)
| | | | | | | neigh_GO_term =
endomembrane_system_: sce03020 (0.0)
| | | \ | - | | | neigh_GO_term = ligase_activity_:
sce03020 (0.0)
| | | | | | | | neigh_GO_term =
translation_regulator_activity_: sce03020 (0.0)
| | | | | | | | neigh_GO_term = peptidase_activity_:
sce03020 (0.0)
| | | | | | | | neigh_GO_term =
microtubule_organizing_center_: sce030̄20(0.0)
| | | | | | | neigh_GO_term =
phosphoprotein_phosphatase_activity_: sce03020 (0.0)
| | | | | | | | neigh GO term =
carbohydrate_metabolic_process_: sce0\overline{3}02\overline{0}}\mathrm{ (0.0)
| | | | | | | | neigh_GO_term = cell_budding_:
sce03020 (0.0)
| | | | | | | | neigh_GO_term =
nuclear_organization_and_biogenesis_:- scē03020 (0.0)
| | | | | \ \ | neig}h_GO_term =
extracellular_region_: sce03020 (0.0)
| | | | | | | | neigh_GO_term =
cytoskeleton_organization_and_biogenesis_: sce03020 (0.0)
| | | | | | | | neigh_GO_term = electron_transport_:
sce03020 (0.0)
| | | | | | | | neigh_GO_term =
cellular_respiration_: sce03020 (0.0)
| | | | | | | start > 412773: sce00790 (35.93/17.95)
    start > 413641: sce03030 (44.66/26.95)
    start > 444317
| | | | start <= 495002
| | | | | | start <= 488387: sce00970 (24.95/8.47)
| | | | | | start > 488387
| | | | | | | neigh_GO_term = nucleus_: sce00620
(5.54/1.91)
| | | | | | | neigh_GO_term = helicase_activity_:
sce00010 (0.0)
| | | | | | | neigh_GO_term = DNA_metabolic_process_:
sce00010 (17.01/4.51)
| | | | | | | neigh_GO_term = vacuole_: sce00010 (0.0)
| | | | | | neigh_GO_term = moleculār_function_:
sce00010 (3.3/1.36)
| | | | | | | neigh_GO_term = biological_process_:
sce00252 (5.65/2.63)
```

```
| | | | | | neigh GO term = cellular component :
sce00010 (0.0)
| | | | | | | neigh_GO_term =
vitamin_metabolic_process_: sce00010 (0.0)
| | | | | | | neigh_GO_term = protein_binding_:
sce00010 (0.0)
| | | | | | | neigh GO term =
oxidoreductase_activity_: sce00010 (0.0)
| | | | | | | neigh_GO_term = cellular_bud_: sce00010
(0.0)
| | | | | | | neigh_GO_term = hydrolase_activity_:
sce00620 (3.45/1.51)
| | | | | | | neigh_GO_term = cytokinesis_: sce00010
(0.0)
| | | | | | | neigh_GO_term =
transcription_regulator_activity_: sce00010 (0.0)
| | | | | | | neigh_GO_term =
protein_modification_process_: sc\overline{e}00\overline{0}10 (0.0)
| | | | | | | neigh_GO_term = membrane_fraction_:
sce00010 (0.0)
| | | | | | | neigh_GO_term = peroxisome_: sce00010
(0.0)
| | | | | | neigh GO term = signal transduction :
sce00010 (0.0)
| | | | | | | neigh_GO_term = mitochondrial_envelope_:
sce00010 (0.0)
| | | | | | | neigh_GO_term =
enzyme_regulator_activity_: sce00010-(0.0)
| | | | \ | | neigh_GO_term = transporter_activity_:
sce00010 (3.54/1.84)
| | | | | | | neigh_GO_term =
organelle_organization_and_biogenesis_: sce00010 (0.0)
| | | | | | | neigh_GO_term = transport_: sce00010
(0.0)
| | | | | | | neigh_GO_term = transferase_activity_:
sce00010 (0.0)
| | | | | | | neigh_GO_term =
membrane_organization_and_biogenesis_: sce00620 (5.04/1.29)
| | | | | | | | neigh_GO_term = membrane_: sce00010
(0.0)
| | | | | | | neigh_GO_term = conjugation_: sce00010
(0.0) | | | | | neigh GO term = cytoplasm : sce00620
(1.05/0.27)
| | | | | | | neigh_GO_term = endoplasmic_reticulum_:
sce00010 (0.0)
| | | | | | | neigh_GO_term = plasma_membrane_:
sce00010 (0.0)
| | | | | | neigh_GO_term = cell_wall_: sce00010
(0.0)
| | | | | | | neigh_GO_term =
structural_molecule_activity_: sce00010 (0.0)
| | | | | | | neigh_GO_term =
cell_wall_organization_and_biogenesis_: sce00010 (0.0)
```



```
| | | | | | | neigh GO term = cytoplasmic membrane-
bound_vesicle_: sce00010 (0.0)
| | | | | | | neigh_GO_term = pseudohyphal_growth_:
sce00010 (0.0)
| | | | | | | neigh_GO_term = endomembrane_system_:
sce00010 (0.0)
| | | | | | neigh GO term = ligase activity :
sce00010 (0.0)
| | | | | | | neigh_GO_term =
translation_regulator_activity_: sce00010 (0.0)
| | | | | | | neigh_GO_term = peptidase_activity_:
sce00010 (0.0)
| | | | | | | neigh_GO_term =
microtubule_organizing_center_: sce000010 (0.0)
| | | | | | | neigh_GO_term =
phosphoprotein_phosphatase_activity_: sce00010 (0.0)
| | | | | | | neigh GO term =
carbohydrate_metabolic_process_ sce\overline{00010 (0.0)}
| | | | | | | neigh_GO_term = cell_budding_: sce00010
(0.0)
| | | | | | | neigh_GO_term =
nuclear_organization_and_biogenesis_: sce00010 (0.0)
| | | | | ` \ neigh_G\overline{O_term = extracellular_region_:}
sce00010 (0.0)
| | | | | | | neigh_GO_term =
cytoskeleton_organization_and_biogenesis_: sce00010 (0.0)
| | | | | | | neigh_GO_term = electron_transport_:
sce00010 (0.0)
| | | | | | neigh_GO_term = cellular_respiration_:
sce00010 (0.0)
        |
        strand = C
        start <= 43074
        | start <= 17248
            start <= 12876: sce00750 (7.13)
            start > 12876: sce00051 (77.04/59.47)
            start > 17248: sce00400 (14.09)
            start > 43074
            start <= 302682
            start <= 155101
                    start <= 57573: sce00562 (59.41/28.84)
                    start > 57573
                        start <= 82806: sce03010 (18.43)
                        start > 82806
                        | start <= 106695: sce04010 (18.52)
                start > 106695
                    | start <= 117349: sce00100 (14.36)
                            start > 117349: sce04010 (14.93)
            start > 155101
                    start <= 197944
                        start <= 182876: sce00230 (43.09/24.01)
                        start > 182876: sce00030 (39.03/16.84)
```

```
| | | | | start > 197944
    | start <= 237664: sce01030 (38.01/16.32)
    start > 237664: sce04070 (37.67/19.01)
    start > 302682
        start <= 380833
        start <= 359598
        start <= 331324: sce00564 (24.65/10.45)
        start > 331324: sce00230 (42.27/22.06)
        start > 359598
        distance <= 333: sce00565 (3.82/0.27)
        distance > 333: sce00564 (34.96/20.54)
        start > 380833
        start <= 430089
        start <= 424497: sce04070 (24.84/10.58)
        start > 424497: sce00620 (26.0/10.97)
        start > 430089
            start <= 440572: sce04010 (19.46)
            start > 440572: sce03010 (19.6)
start > 529943
| start <= 678801
    strand = W
        start <= 559003
            start <= 538174
            start <= 531726: sce00190 (17.31)
            start > 531726: sce01030 (45.77/20.2)
        start > 538174
            start <= 542964: sce00590 (21.54)
            start > 542964: sce01031 (38.98/16.68)
        start > 559003
        start <= 640398
        start <= 627144: sce00720 (66.28/38.98)
        start > 627144: sce00561 (22.81)
        start > 640398
            start <= 674925
            start <= 665341
            | start <= 647434: sce00240 (13.47)
            | start > 647434: sce00120 (18.38)
            start > 665341: sce00750 (15.52)
            start > 674925: sce00510 (45.79/20.29)
        strand = C
        start <= 631063
        start <= 627144
            start <= 559003: sce00020 (17.03)
            start > 559003: sce03010 (14.26)
        start > 627144: sce00020 (22.1/9.23)
        start > 631063
        start <= 635299: sce00230 (41.24/24.68)
        start > 635299
            start <= 647434: sce04140 (17.94)
            start > 647434: sce00620 (28.14/11.15)
    start > 678801
        start <= 690323
            strand = W
        | start <= 687637: sce00790 (18.83)
```





## Decision Tree Generated for Chromosome Fifteen:

```
start <= 525279
| start <= 263476
| | strand = C
| | | start <= 137825
    | start <= 82920
        | start <= 70325
        | | start <= 58759
        | | start <= 24293: sce00500 (14.0)
        | start > 24293: sce00740 (25.15)
            start > 58759: sce00051 (15.64)
        start > 70325: sce00020 (75.42/57.15)
        start > 82920
        start <= 101475: sce03010 (40.54)
        start > 101475
            start <= 115808: sce04010 (8.09)
            start > 115808: sce00970 (38.31/18.21)
        start > 137825
        start <= 160594
        start <= 139045: sce00130 (12.76)
        start > 139045: sce00071 (88.25/72.05)
        start > 160594
            start <= 212244
            | start <= 205885: sce00562 (32.39/12.48)
            start > 205885: sce00920 (19.02)
            start > 212244
            start <= 226074: sce00730 (19.99)
            start > 226074
            | start <= 238618: sce00220 (31.58/15.73)
                        start > 238618: sce03010 (16.64)
        strand = W
        start <= 226074
            start <= 185438
        | start <= 70325
        start <= 58759: sce00220 (18.44)
```

```
| | | | start > 58759: sce04111 (36.65/18.18)
start > 70325
    | start <= 139045
    | | start <= 82920: sce03010 (12.08)
    | start > 82920: sce03030 (10.63)
    start > 139045: sce00190 (16.2)
start > 185438
    start <= 217126
    | start <= 212244: sce00230 (30.73/11.73)
    | start > 212244: sce00564 (20.87)
    start > 217126: sce00220 (44.68/25.58)
start > 226074
    start <= 254296
    start <= 238618: sce00480 (41.84/20.76)
    start > 238618: sce03010 (14.76)
    start > 254296
    start <= 255336: sce03050 (12.22)
    start > 255336: sce00970 (49.43/26.71)
start > 263476
    strand = C
    start <= 415908
start <= 325250
    start <= 303036
        | start <= 292075: sce04130 (19.35)
        start > 292075: sce04120 (15.92)
        start > 303036: sce00240 (46.53/29.3)
        start > 325250
        start <= 342415: sce04010 (7.14)
        start > 342415
            neigh_num <= 1: sce00790 (9.41/3.73)
            neigh_num > 1: sce00500 (21.28/6.34)
        start > 415908
        start <= 482035
        start <= 453464: sce00510 (37.94/16.3)
        start > 453464: sce00240 (31.98/13.18)
        start > 482035
        start <= 505795: sce00030 (44.7/21.91)
        start > 505795
        | neigh strand = W: sce00510 (22.35/6.73)
    | neigh_strand = C: sce01030 (8.44/2.54)
strand = W
        start <= 379781
        start <= 342415
            start <= 325250: sce04111 (21.5)
            start > 325250: sce01030 (29.4/9.49)
        start > 342415: sce04111 (23.7)
        start > 379781
        start <= 519122
            start <= 415908
            | start <= 400348: sce04130 (11.81)
            start > 400348: sce00620 (21.57)
            start > 415908
            start <= 453464
            | start <= 444688: sce03010 (20.55)
```




```
| | | | | | neigh_GO_term = conjugation_: sce00380 (0.0)
| | | | | | neigh_GO_term =
structural_molecule_activity_: sce00380 (0.0)
| | | | | | neigh_GO_term = vacuole_: sce00380 (0.0)
```



```
| | | | | | - neigh_GO_term = cytokinesis_: sce00380 (0.0)
| | | | | | neigh_GO_term = cellular_bud_: sce00380
(0.0)
| | | | | | neigh_GO_term = site_of_polarized_growth_:
sce00380 (0.0)
| | | | | | neigh_GO_term =
cytoskeleton_organization_and_biogenesis_: sce00380 (0.0)
| | | | | | neigh_GO_term = endoplasmic_reticulum_:
sce00380 (0.0)
| | | | | | neigh_GO_term = signal_transduction_:
sce00380 (0.0)
| | | | | | neigh_GO_term = Golgi_apparatus_: sce00380
(0.0)
| | | | | | neigh_GO_term = lipid_metabolic_process_:
sce00380 (0.0)
| | | | | | neigh_GO_term = membrane_fraction_: sce00380
(0.0)
| | | | | | neigh_GO_term = signal_transducer_activity_:
sce00380 (0.0)
| | | | | | neigh_GO_term = meiosis_: sce00350
(6.48/4.11)
| | | | | | neigh_GO_term = ligase_activity_: sce00380
(0.0)
| | | | | | neigh_GO_term = helicase_activity_: sce00380
(0.0)
| | | | | | neigh_GO_term =
microtubule_organizing_center__: sce00380 (0.0)
| | | | | | neigh_GO_term =
generation_of_precursor_metabolites_and_energy_: sce00380 (0.0)
| | | | | | neigh_GO_term = protein_binding_: sce00380
(0.0)
| | | | | | neigh_GO_term = endomembrane_system_:
sce00340 (4.62/2.75)
| | | | | | neigh_GO_term = vesicle-mediated_transport_:
sce00380 (0.0)
| | | | | | neigh_GO_term = lyase_activity_: sce00440
(5.5/2.98)
| | | | | | neigh_GO_term = peptidase_activity_:
sce00380 (0.0)
| | | | | | neigh_GO_term = protein_catabolic_process_:
sce00380 (0.0)
| | | | | | neigh_GO_term = mitochondrial_envelope_:
sce00380 (0.0)
| | | | | | neigh_GO_term = cellular_respiration_:
sce00380 (0.0)
| | | | | | neigh_GO_term =
phosphoprotein_phosphatase_activity_: sce00380 (0.0)
```





```
| | | | | | | neigh GO term =
site_of_polarized_growth_: sce001\overline{2}0\mathrm{ (0.0)}
| \ | | | | ` neigh_GO_term =
cytoskeleton_organization_and_biogenesis_: sce00120 (0.0)
| | | | | | | neigh_GO_term = endoplasmic_reticulum_:
sce00120 (0.0)
| | | | | | neigh_GO_term = signal_transduction_:
sce00120 (0.0)
| | | | | | | neigh_GO_term = Golgi_apparatus_:
sce00120 (0.0)
| | | | | | | neigh_GO_term =
lipid_metabolic_process_: sce0012\overline{0}
| | | | | | - | neigh_GO_term = membrane_fraction_:
sce00120 (0.0)
| | | | | | | neigh_GO_term =
signal_transducer_activity_: sce00120 (0.0)
| | | | | | | neigh_GO_term = meiosis_: sce00120 (0.0)
| | | | | | neigh_GO_term = ligase_äctivity_:
sce00120 (0.0)
| | | | | | | neigh_GO_term = helicase_activity_:
sce00120 (0.0)
| | | | | | | neigh_GO_term =
microtubule_organizing_center_: s\overline{ce0}00120 (0.0)
| | | | | | | neigh_GO_term =
generation_of_precursor_metabolites_and_energy_: sce00120 (0.0)
| | | | | | | neigh_GO_term = protein_binding_:
sce00120 (0.0)
| | | | | | | neigh_GO_term = endomembrane_system_:
sce00120 (0.0)
| | | | | | | neigh_GO_term = vesicle-
mediated_transport_: sce00120 (0.\overline{0})
| | | | | | neigh_GO_term = lyase_activity_:
sce00120 (0.0)
| | | | | | | neigh_GO_term = peptidase_activity_:
sce00120 (0.0)
| | | | | | | neigh_GO_term =
protein_catabolic_process_: sce00120 (0.0)
| | | | | | | neigh_GO_term = mitochondrial_envelope_:
sce00120 (0.0)
| | | | | | | neigh_GO_term = cellular_respiration_:
sce00120 (0.0)
| | | | | | | neigh_GO_term =
phosphoprotein_phosphatase_activity_: sce00120 (0.0)
| | | | | | | neigh_GO_term = pseudohyphal_growth_:
sce00260 (5.68/3.56)
| | | | | | | neigh_GO_term =
membrane_organization_and_biogenesis_: sce00120 (0.0)
| | | | | | | neigh_GO_term = cytoskeleton_: sce00120
(0.0)
| | | | | | neigh GO term = electron transport :
sce00120 (0.0)
| | | | | | neigh_GO_term = cellular_homeostasis_:
sce00120 (0.0)
```






## Decision Tree Generated for Chromosome Sixteen:

```
start <= 568993
| start <= 423889
| | start <= 179276
| | | strand = W
| | | | start <= 103232
| | | | | start <= 66614
| | | | | start <= 35236
| | | | | | start <= 30079: sce00190 (18.68)
| | | | | | | | | start > | 300
| | | | | | neigh_num<= 1: sce00720 (7.31/3.3)
| | | | | | neigh_num > 1
(4.58/0.43)
| | | | | | | | distance > 1808: sce00720
(7.22/1.75)
| | | | | start > 66614
    | start > 66614
    | start > 88517: sce00500 (35.77/17.09)
    start > 103232
    start <= 121167
    | start <= 107275: sce04130 (17.67)
    | start > 107275: sce01040 (55.67/17.86)
    start > 121167
    | start <= 173151: sce03010 (26.69)
```










```
| | | | | | neigh GO term = cytokinesis : sce00620 (0.0)
| | | | | | neigh_GO_term = response_to_stress_:
sce00620 (0.0)
| | | | | | neigh_GO_term =
translation_regulator_activity_: sce00620 (0.0)
| | | | | | neigh_GO_term = RNA_metabolic_process_:
sce00071 (15.15/11.1)
| | | | | | neigh_GO_term = vacuole_: sce00620 (0.0)
| | | | | | neigh_GO_term = protein_kinase_activity_:
sce00620 (13.99/9.31)
| | | | | | neigh_GO_term =
organelle_organization_and_biōge\overline{nesis_: sce00620 (0.0)}
| | | | | | neígh_GO_term = chromosome_: sce00620 (0.0)
| | | | | | neigh_GO_term = cytoskeleton_: sce00620
(0.0)
| | | | | | neigh_GO_term = endomembrane_system_:
sce00620 (0.0)
| | | | | | neigh_GO_term = endoplasmic_reticulum_:
sce00620 (0.0)
| | | | | | neigh_GO_term =
protein_modification_process_: sce00310 (10.47/5.2)
| | | | | | neigh_GO_term =
membrane organization and biog
| | \ | | | neigh_GO_term = lipid_metabolic_process_:
sce00620 (0.0)
| | | | | | neigh_GO_term = cellular_homeostasis_:
sce00620 (0.0)
| | | | | | neigh_GO_term =
cell_wall_organization_and_biōgen̄esis_: sce00650 (10.27/7.2)
| | | | | | neigh_GO_term = isomerase_activity_:
sce00620 (0.0)
| | | | | | neigh_GO_term = meiosis_: sce00620 (0.0)
| | | | | | neigh_GO_term =
nuclear_organization_and_biogene\overline{sis_: sce00620 (0.0)}
| | | | | | neigh_GO_term = extracellular_region_:
sce00620 (0.0)
| | | | | | neigh_GO_term =
phosphoprotein_phosphatase_activitty_: sce00620 (0.0)
| | | | | | neigh_GO_term = protein_catabolic_process_:
sce00620 (0.0)
| | | | | | neigh_GO_term =
nucleotidyltransferase_activity_: sce00620 (0.0)
| | | | | | neigh_GO_term = ligase_activity_: sce00620
(0.0)
| | | | | | neigh_GO_term = cell_wall_: sce00620 (0.0)
| | | | | | neigh_GO_term = sporulatiōn_: sce00650
(12.59/9.23)
| | | | neigh_GO_term = peroxisome_: sce00620 (0.0)
| | | | | | neigh_GO_term = mitochondríal_envelope_:
sce00620 (0.0)
| | | | | | neigh_GO_term =
anatomical_structure_morphogenesis_: sce00620 (0.0)
| | | | | | neigh_GO_term = signal_transducer_activity_:
sce00620 (0.0)
```



```
| | | | | | | | neigh GO term = membrane : sce00562
(0.0)
| | | | | | | neigh_GO_term = hydrolase_activity_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term =
cellular_respiration_: sce00562 (0.0)
| | \ | | \ | | neigh_GO_term = cell_cortex_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term =
cytoskeleton_organization_and_biogenesis_: sce00562 (0.0)
| | | | | | | | neigh_GO_term =
pseudohyphal_growth_: sce00562 (0.0)
| | | | | | | | neigh_GO_term =
signal_transduction_: sce00562 (0.0)
| | | | | | | | neigh_GO_term = cell_cycle_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term = nucleolus_: sce00562
(0.0)
| | | | | | | | neigh_GO_term =
ribosome_biogenesis_and_assembly_: sce00562 (0.0)
| | | | | | | | neigh_GO_term =
carbohydrate_metabolic_process_: sce00562 (0.0)
| | | | | | | | _ neigh_GO_term = Golgi_apparatus_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term = vesicle-
mediated_transport_: sce00562 (0.0)
| | | | | | | | neigh_GO_term =
vitamin_metabolic_process_: sce00562 (0.0
| | | | | | | | | neigh_GO_term = translation_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term =
enzyme_regulator_activity_: sce00562 (0.0)
| | | | | | | neigh_GO_term = conjugation_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term =
microtubule_organizing_center_: sce00562 (0.0)
| | | | | | | neigh_GO_term =
structural_molecule_activity_: sce00562 (0.0)
| | | | | - | | \ neigh_GO_term = motor_activity_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term =
transcription_regulator_activity_: sce00562 (0.0)
| | | | | | | | neigh_GO_term = transcription_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term =
oxidoreductase_activity_: sce00562 (0.0)
| | | | | | | | neigh_GO_term = ribosome_: sce00562
(0.0)
| | | | | | | | neigh_GO_term = RNA_binding_:
sce00562 (0.0)
| | | | | | | neigh_GO_term = DNA_binding_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term = cellular_bud_:
sce00562 (0.0)
```

```
| | | | | | | | neigh GO term =
site_of_polarized_growth_: sce00562 (\overline{0.0)}
| | | | | | | | neigh_GO_term = cytokinesis_:
sce00562 (0.0)
| | | | | | | neigh_GO_term = response_to_stress_:
sce00562 (0.0)
| | | | | | | | neigh GO term =
translation_regulator_activity_: sce0\overline{0}56\overline{2}}\mathrm{ (0.0)
| | | | | | | | neigh_GO_term =
RNA_metabolic_process_: sce00562 (0.0)
| | | | | | | | neigh_GO_term = vacuole_: sce00562
(0.0)
| | | | | | | | neigh_GO_term =
protein_kinase_activity_: sce00562 (0.0)
| | | | | | | | neigh_GO_term =
organelle_organization_and_biogenesis_: sce00562 (0.0)
| | | | | | | | neigh_GO_term = chromosome_:
sce00562 (0.0)
| | | | | | | neigh_GO_term = cytoskeleton_:
sce00562 (0.0)
| | | | | | | neigh_GO_term =
endomembrane_system_: sce00562 (0.0)
| | | | | - | | | neigh_GO_term =
endoplasmic_reticulum_: sce00562 (0.0)
| | | | | | | | neigh_GO_term =
protein_modification_process_: sce00562 (0.0)
| | | | | | | | neigh_GO_term =
membrane_organization_and_biogenesis_: sce00562 (0.0)
| | \ | | | | | | neigh̆GO_term =
lipid_metabolic_process_: sce00562 (0.0)
| | | | | | | | neigh_GO_term =
cellular_homeostasis_: sce00562 (0.0)
| | | | | | | | neigh_GO_term =
cell_wall_organization_and_biogenesis_: sce00562 (0.0)
| | | | | | | | neigh_GO_term = isomerase_activity_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term = meiosis_: sce00562
(0.0)
| | | | | | | | neigh GO term =
nuclear_organization_and_biogenesis_: sce00562 (0.0)
| | | | | | | | neigh_GO_term =
extracellular_region_: sce00562 (0.0)
| | | | | | | | neigh_GO_term =
phosphoprotein_phosphatase_activity_: sce00562 (0.0)
| | | | | | | | neiğh_GO_term =
protein_catabolic_process_: sce00562 (0.0)
| | | | | | | | neigh_GO_term =
nucleotidyltransferase_activity_: sce00562 (0.0)
| | | | | | | | neigh_GO_term = ligase_activity_:
sce00562 (0.0)
| | | | | | | | neigh_GO_term = cell_wall_: sce00562
(0.0)
| | | | | | | neigh_GO_term = sporulation_:
sce00562 (0.0)
```








```
| | | | | neigh_GO_term = electron_transport_: sce00240
| | | | start > 599867
| | | | | start <= 643833: sce00340 (38.79/17.0)
| | | | | start > 643833
|
    | start <= 664957: sce04140 (12.86)
    | start > 664957: sce00400 (21.29)
    start > 677162
    | start <= 692415
    start <= 685434: sce00220 (61.23/29.33)
    start > 685434: sce00361 (23.96/11.98)
        start > 692415
        | start <= 694835: sce00710 (33.25/15.16)
        start > 694835: sce00260 (37.31/14.79)
        start > 703968
        start <= 732347
        | start <= 713026: sce03060 (17.23)
        start > 713026: sce03010 (20.47)
        start > 732347
        | start <= 868300
            start <= 752255: sce03020 (37.99/22.07)
            start > 752255: sce04111 (29.76)
        start > 868300
        | start <= 895957: sce00920 (17.29)
        start > 895957: sce03020 (61.15/38.61)
```


## Appendix I. Decision Trees Generated for Classifying Features to Gene Ontology Terms

In the tree structure, a colon introduces the name of the Gene Ontology term that has been assigned to a particular leaf. The Gene Ontology Term is followed by the number of instances that reach that leaf. The number is expressed as a decimal because of the way the classification algorithm uses fractions to handle missing values. The number of incorrectly classified instances that reach that leaf, if they exist, are reported after the "/". For example, if the numbers (2.0/1.0) are reported, it means that two instances reached the leaf, of which one is incorrectly classified.

## Decision Tree Generated for Chromosome One:

## For Cellular Component










## For Biological Process

```
| start <= 52596
| | start <= 33449
| | | start <= 9017
| | | | strand = C
    start <= 2480: sporulation_(6.68)
    start > 2480: transport (\overline{15.27)}
        strand = W: biological_procèss_ (8.64)
        start > 9017
        | start <= 27969: biological_process_ (69.72)
        start > 27969
    start <= 31568:
amino_acid_and_derivative_metabolic_process_ (17.95)
| | | | | start > 31568: b̄iological_process_ (14.07)
| | start > 33449
```






## Decision Tree Generated for Chromosome Two:

## For Cellular Component

start <= 496863
| | start <= 227639





| | | | | | | | | | | NeighGO_term = signal_transduction_: nucleus_ (0.0) | | | | | | | | | | | generation_of_precursor_metabolites_and_energy_: nucleus_ (0.0)
 vitamin_metabolic_process_: nucleus_(0.0)
$\left|{ }^{-}\right| \quad|\quad|^{-}| |^{-}|\quad|{ }^{-}|\quad| \quad$ NeighGo_term $=$ nucleotidyltransferase_activity_: nucleus_ (0.0)
| | | | | | | | | | | | NeighGo_term = cytoskeleton_: nucleus_ (0.0)
 cell_wall_: nucleus_ (0.0)
| $\mid$ | | | | | | | | Golgi_apparatus_: nucleus_(0.0) $\mid$ peroxisome_: nucleus_(0.0)
| | | | | $\mid$ | | | | | NeighGo_term = ligase_activity_: nucleus_(0.0)
| | | | | | | | | | ${ }^{-}$| NeighGo_term = carbohydrate_metabolic_process_: nucleus_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ | $\mid$ NeighGo_term $=$ lyase_activity_: nucleus_(0.0)
| ${ }^{-}| |^{-} \mid$| $\mid$| | | pseudohyphal_growth_: nucleus_(0.0)
$|\quad| \quad|\quad| \quad-|\quad| \quad \mid \quad$ | $\mid$ NeighGo_term $=$ anatomical_structure_morphogenesis_: nucleus_ (0.0)
$\left|\left|\left.\right|^{-}\right|\right.$| $|$| | | $\mid$| $\mid$NeighGo_term = cellular_homeostasis_ nucleus_(0.0)
| | $\quad|\quad| \quad \overline{\mid}\left|{ }^{-}\right|$start > 117592 | | | | | | | | | start <= 121877: cellular_component_(29.18/11.67)
| | | | | | | | | | start > 121877: cytoplasm_ (24.52)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>124762$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | strand $=C$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ start $<=132427$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\|\quad\| \quad$ distance $<=3982:$ |

cellular bud (12.31/3.05)
$|\quad| \quad|\quad| \quad|\quad| \quad$ distance > 3982:
site_of_polarized_growth_(13.46/2.14)
| $\overline{\text { - }} \mid$ | | $\mid$ | $\mid$ | start > 132427:
mitochondrion_(12.95)
| | | | | | | | | strand = W
| | | | | | | | start <= 125128:
membrane_fraction (22.82)
$\begin{array}{llllllllll}\mid & \mid & \overline{ } & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }>125128 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=126831:\end{array}$
cellular_component_ (11.11)
$|\quad| \quad\left|\quad{ }^{-}\right| \quad|\quad| \quad|\quad| \quad \mid \quad$ start > 126831: nucleus (15.12)
| | | | | start > 135519
| | | | | | start <= 151223
$|\quad| \quad|\quad| \quad \mid \quad$ strand $=\mathrm{C}$







```
| | | | | | | | | | | | | | NeighGO_term
= membrane_: chromosome_(0.0)
| | | | | | | | | | | | | | NeighGO_term
= nuclear_organization_and_biogenesis_: chromosome_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= membrane_organization_and_biogenesis_: chromosome_(0.0)
| | | | | | - | - | | | - | | | - | NeighGO term
= organelle_organization_and_biogenesis_: chromosome_ (0.0)
| | | | | | \ | | | | | | | NeighGO_term
= nucleolus_: chromosome_(0.0)
| | | | | | | | | | | | | | NeighGO_term
= RNA_binding_: chromosome_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= RNA_metabolic_process_: chromosome_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= ribosome_biogenesis_and_assembly_: chromosome_ (0.0)
| | | | | | | | | | | | - | | NeighGo_term
= peptidase_activity_: chromosome_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= microtubule_organizing_center_: chromosome_ (0.0) | | | | | | | | | | | | | | | | NeighGO_term
= motor_activity_: chromosome_ (0.0)
| | | | \ | | | | | | | | | | NeighGO term
= cytoskeleton_organization_and_biogenesis_: nucleus_(2.93/1.8\overline{5})
| | | | | | | | | | | | | | NeighGO_term
= cellular_bud_: chromosome_(0.0)
| | | | | | | | | | | | | | NeighGO_term
= site_of_polarized_growth_: chromosome_ (0.0)
| | - | | | - | | - | | | - | | | | NeighGO_term
= enzyme_regulator_activity_: chromosome_ (0.0)
| | \ | | | | - | | | \ | | | NeighGO_term
= cytokinesis_: chromosome_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= hydrolase_activity_: chromosome_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= phosphoprotein_phosphatase_activity_: chromosome_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= protein_modification_process_: chromosome_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= transferase_activity_: chromosome_(4.76/2.63)
| | | | | | | | | | | | | | NeighGO_term
= DNA_metabolic_process_: chromosome_(0.0)
| | | | | | | | | | | | | | NeighGO_term
= membrane_fraction_: chromosome_ (0.0)
| | | | | - | | | \ | | | | | NeighGO_term
= transport_: chromosome_(2.22/0.86)
| | | | | | | | | | | | | | NeighGO_term
= oxidoreductase_activity_: chromosome_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= cellular_respiration_: chromosome_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= plasma_membrane_: chromosome_(0.0)
| | | | | | | | | | | | | | NeighGO_term
= transporter_activity_: chromosome_ (0.0)
```



| | | | | | | | | | | NeighGO_term = molecular_function_: mitochondrion_(12.05/5.15)

```
biological_process_: plasma_membrane_ (0.0) | | | | | | | | | | NeighGO_term=
```

endomembrane_system_: plasma_membrane_ (0.0)
| | | $\mid$ | - | $\mid$ | ${ }^{-}|\quad| \quad$ NeighGo_term $=$
vesicle-mediated_transport_: plasma_membrane_ (0.0)
| | | | | | | | | | | | NeighGo_term =
cell_wall_organization_and_biogenesis_: plasma_membrane_(3.66/1.58)

cellular_component_: membrane_fraction_(2.94/1.28)
| | $\mid$ | ${ }^{-} \mid$| $\mid$| ${ }^{-}$| NeighGo_term $=$
ribosome_: plasma_membrane_(7.04/2.47)
| | | | | | | | | | | | NeighGO_term =
structural_molecule_activity_: plasma_membrane_ (0.0)
| | | | | | | | | | | | NeighGo_term =
translation_: plasma_membrane_(2.94/1.08)
| | | - | $\mid$ | | | | |
lipid_metabolic_process_: plasma_membrane_ (0.0)
mitochondrion_: plasma_membrane_(0.0)
| | | ${ }^{-}| |^{-}|\quad|-|\quad| \quad$ | NeighGo_term $=$
nucleus_: plasma_membrane_ (0.0)
| | | | | | | | | | NeighGO_term =
protein_kinase_activity_: plasma_membrane_ (0.0)

response_to_stress_: plasma_membrane_ (0.0)
| | $\quad$ | | $\quad$ | $\mid$ | | $\mid$ | NeighGO_term =
protein_binding_: plasma_membrane_ (0.0)
cell_cycle_: plasma_membrane_(0.0)
| $\mid$ | | | | | $\mid$ | | | | NeighGo_term $=$
membrane_: plasma_membrane_ (0.0)
| | $\mid$ | | | | $\quad$ | | |
nuclear_organization_and_biogenesis_: plasma_membrane_ (0.0)

membrane_organization_and_biogenesis_: plasma_membrane_(0.0)
$|\quad| \quad|\quad| \quad|\quad|^{-}|\quad| \quad|\quad| \quad{ }^{-}$NeighG $\bar{O}$ term $=$
organelle_organization_and_biogenesis_: plasma_membrane_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGo_term $=$
nucleolus_: plasma_membrane_ (0.0)

RNA_binding_: plasma_membrane_ (0.0)
| $\mid$ | $\mid$ | $\mid$ | | | $\mid$ NeighGo_term $=$
RNA_metabolic_process_: plasma_membrane_ (0.0)
| | | | | | | | | | | NeighGO_term =
ribosome_biogenesis_and_assembly_: plasma_membrane_ (0. $\overline{0}$ )

peptidase_activity_: plasma_membrane_(0.0)
| | | | | | | - | $\mid$ | | NeighGo_term =
microtubule_organizing_center_: plasma_membrane_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGo_term $=$
motor_activity_: plasma_membrane_ (0.0)

| | | | | | | | | | | NeighGO_term = mitochondrial_envelope_: plasma_membrane_ (0.0) | transcription_: plasma_membrane_ (0.0)
| | | | | | | | | | | NeighGO_term = translation_regulator_activity_: plasma_membrane_(2.54/1.44)
 conjugation_: plasma_membrane_(0.0)
| | | | | | | | | | | | NeighGo_term = cytoplasmic_membrane-bound_vesicle_: plasma_membrane_( $\overline{0} .0)$
 sporulation_: plasma_membrane_(0.0)
| | | | | $\mid$ | | | | | | NeighGo_term = cell_budding_: plasma_membrane_(0.0)
| | | | | | | | | | | NeighGo_term = vacuole_: plasma_membrane_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGo term $=$ signal_transduction_: membrane fraction_(4.73/2.53)
| | | | | | | | | | | | NeighGO_term = generation_of_precursor_metabolites_and_energy_: plasma_membrane_ (0.0)
| | | | | | | | | | | $\mid$ NeighGO_term = vitamin_metabolic_process_: plasma_membrane_ (0.0)
| | $\mid$ | | | ${ }^{-}$| $\left.\right|^{-}|\quad|{ }^{-}$| NeighGo_term = nucleotidyltransferase_activity_: plasma_membrane_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ | $\mid$ ighGo_term $=$ cytoskeleton_: plasma_membrane_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGo_term $=$ cell_wall_: plasma_membrane_(0.0)
| | | | | | | | | | | | NeighGo_term = Golgi_apparatus_: plasma_membrane_(0.0) NeighGO term= peroxisome_: plasma_membrane_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ | $\mid$ NeighGo_term $=$ ligase_activity_: plasma_membrane_ (0.0)
| | | | | | $\mid$ | | | | | NeighGo_term = carbohydrate_metabolic_process_: plasma_membrane_ (0.0)
 lyase_activity_: plasma_membrane_(0.0)
| | ${ }^{-}$| | | $\mid$| $\mid$| | NeighGo_term = pseudohyphal_growth_: plasma_membrane_ (0.0)
| | | | | | | | | | | | NeighGo_term = anatomical_structure_morphogenesis_: plasma_membrane_( $\overline{0} .0)$
| | ${ }^{-}|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ | Neigh̄Go_term $=$ cellular_homeostasis_: plasma_membrane_ (0.0)
| | $\quad$ | | $\quad|\quad| \quad|\quad|-$ start $>393118$
| | | | | | | | | | start <= 398271:
ribosome_(36.63/17.05)
| | | | | | | | | | | start > 398271
$\mid$ | | | | | | | | | | | $\mid$ strand $=C$ | | | | | | | | | | | | start <= 400884: cytoplasm_(12.85)
mitochondrion_(21.65)





 (14.2)









## For Molecular Function

| start $<=147212$
$\mid$
$\mid$









```
| | | | | | | | | | | | | | | start <=
393118: molecular_function_(36.11)
| | | | | | | | | | | | | | | start >
393118
| | | | | | | | | | | | | | | | |
start <= 398271: translation_regulator_activity_ (18.45)
| | | | | | | \ | | - | | - | | | |
start > 398271: hydrolase_activity_ (14.16)
| | | | | | | | | | | | | | start >
400884
| | | | | | | | | | | | | | | start <=
4 1 5 9 7 7
start <= 411048
| | | | | | | | | | | | | | | | |
start <= 405954: structural_molecule_activity_ (11.89)
| | | | | | | | | | | | | | | | | |
start > 405954: molecular function_ (14.68)
| | | | | | | | | | | | | | | |
start > 411048: structural_molecule_activity_ (12.69)
| | | | | | | | | | | | | | | start >
4 1 5 9 7 7
| | | | | | | | | | | | | | | |
start <= 423759: molecular_function_ (29.92)
| | | | | | | | | | | | | | | |
start > 423759
| | | | | | | | | | | | | | | | | | |
start <= 425760: enzyme_regulator_activity_ (13.35)
| | | | | | | | | | | | | | | | | | |
start > 425760: molecular_function_ (18.6)
| | | | | | | | | | | | | start > 427052
| | | | | | | | | | | | | start <=
427478: transporter_activity_ (11.62)
| | | | | - | | \ | | | | | | start >
427478: hydrolase_activity_ (33.58)
|
4 4 7 7 0 3
| | | | | | | | | | | | | | start <=
444687: molecular_function_(17.52)
| | | | | | | | | | | | | | | start >
444687: protein_binding_ (10.07)
| | | | | | | | | | | | | | start >
447703: molecular_function_(50.64)
| | | | | | | | | | | | | | start > 454816:
protein_binding_ (14.25)
| | | | | | | | | strand = W
| | | | | | | | | | start <= 407163
| | | | | | | | | | start <= 382027
| | | | | | | | | | | | | start <= 380408:
molecular_function_(17.82)
```






| | | | | | | | | | | | start $<=666533$ : oxidoreductase_activity_ (14.05) | | | | | | | | | | | | | start > 666533: molecular_function_(13.73)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>668346$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | 670622: hydrolase_activity_ (10.67) | | | | | | | | | | |

| | | | | | | | | | | | | start > 675308:
hydrolase activity (16.98)
 | | | | | | | | | | | | start <= 620867: protein_binding_ (23.16)
| | | | | | | | | | | | | | start >
620867
| | | | | | | | | | | | | | | start <=
622978: molecular_function_ (18.97)
| | | | | | | | | | | | |
622978: protein_binding_ (17.63)
| | | | | | | | | | | | start > 623572
| | | | | | | | | | | | $\mid$ start <= 625767:
transcription_regulator_activity_ (15.6)

molecular_function_(19.09)
| | | | | | | | | | | start > 629163
| | | | | | | | | | | start <= 633617:
transferase_activity_ (17.17)
| | | | | | | | | | | start > 633617:
transporter_activity_ (18.46)
| | | | | | | | | | start > 635141
$\begin{array}{lllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=650363 \\ \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=645545:\end{array}$
molecular function_(26.57)
| | | $\left.\left.\right|^{-}\right|^{-}$| | | | | | | start > 645545

= W: translation_regulator_activity_ (15.79/7.04)


```
| | | | | | | | | | | | start > 711586:
RNA_binding_(19.07)
| | | | | | | | | | | strand = W:
molecular_function_(52.5/17.0)
| | | | start > 714450
| | | | | | start <= 735525
(41.02)
| | | | | | | start > 719028
```



```
enzyme_regulator_activity_ (21.33)
| | | | | | | | | start > 724451:
molecular_function_(7.97)
| | | | | | | | start > 727069
| | | | | | | | start <= 730382:
transferase activity (29.23)
| | | - | \ | | | start > 730382
| | | | | | | | | start <= 734634:
enzyme_regulator_activity_ (20.08)
| | | | | | | | | | start > 734634:
transferase activity (20.66)
| | | | | \ start > 735525
| | | | | | | start <= 749589: molecular_function_
(89.99)
```



```
(23.17)
| | | | | | | | start > 719028
| | | | | | | | start <= 721385:
structural_molecule_activity_ (30.76)
| | | | | | | | | start > 721385
| | | | | | | | | | start <= 722606:
hydrolase_activity_ (23.49)
| | | | | | | | | | start > 722606:
structural molecule activity (18.03)
| | | | | | | start > 723265
| | | | | | | start <= 724451: molecular_function_
(20.49)
| | | | | | | | start > 724451
| | | | | | | | | start <= 728880:
hydrolase_activity_ (16.17)
| | | | | | | | | start > 728880:
molecular_function_(12.49)
```




```
| | | | | | | | start <= 800517:
molecular_function_(14.96)
| | | | | | | | | start > 800517:
transporter_activity_ (11.74)
| | | | | | | | start > 804469: molecular_function_
(18.53)
| | | | | | | strand = W
| | | | | | | start <= 800517
| | | | | | | | | Neigh_GO_aspect = C:
transcription_regulator_activity_ (24.12/6.86)
| | | | | | | | | Neigh_GO_aspect = F:
DNA_binding_(18.52/3.81)
| | | | | | | | | Neigh_GO_aspect = P:
transcription_regulator_activity_ (17.96/音.79}
| | | | | | | | | start > 800517 
hydrolase_activity_ (14.89)
| | | | | | | | | start > 805345:
molecular_function_(12.17)
```


## For Biological Process

```
| start <= 744847
| | start <= 454524
| | | start <= 398271
| | | | start <= 166137
| | | | | start <= 132427
| | | | | | start <= 48822
| | | | | | strand = C | | | start <= 13879
| | | | | | | | start <= 10551:
biological_process_ (57.19)
| | | | | | | | | start > 10551: vesicle-
mediated_transport_ (19.2)
| | | | | | | | start > 13879
| | | | | | | | start <= 36395
| | | | | | | | | start <= 17696
| | | | | | | | | | | NeighGO_term =
cytoplasm_: signal_transduction_(1.56/0.62)
| | | | | | | | | | | NeighGO_term =
molecular_function_: cell_wall_organization_and_biogenesis_
(13.12/4.4}8
| | | | | | | | | | NeighGO_term =
biological_process_: signal_transduction_(5.9/1.38)
| | | | | | | | | | | NeighGO_term =
endomembrane_system_: signal_transduction_(1.56/0.\overline{3}2)
| | | \ | - | | \ | | | ' NeighGO term = vesicle-
mediated_transport_: cell_wall_organization_and_biog}enesi
(8.33/2.91)
| | | | | | | | | | | NeighGO_term =
cell_wall_organization_and_biogenesis_: signal_transduction_(0.0)
| | | | | | | | | | | NeighGO_term =
cellular_component_: signal_transduction_(0.0)
```

```
| | | | | | | | | | | NeighGO term =
ribosome : signal transduction (0.0)
| | | | | | | | | | | NeighGO_term =
structural_molecule_activity_: signal_transduction_(0.0)
| | | | | | | | | | | NeighGO_term =
translation_: signal_transduction_ (3.49/0.55)
| | | | | \ | | | | | NeighGO_term =
lipid_metabolic_process_: signal_transduction_ (0.0)
| | | | | | | | | | | NeighGO_term =
mitochondrion_: signal_transduction_(0.0)
| | | | | | | | | | NeighGO_term = nucleus_:
signal_transduction_ (0.0)
| | | | | | | | | | | NeighGO term =
protein_kinase_activity_: signal_transduction_(0.0)
| | | | | | | | | | | NeighGO_term =
response_to_stress_: signal_transduction_(0.0)
| | \ | | | | - | | | \ NeighGO_term =
protein_binding_: signal_transduction_ (0.0)
| | | | | | | | | | | NeighGO_term =
cell_cycle_: signal_transduction_ (0.0)
| | | | | | | | | | | NeighGO_term =
membrane_: signal_transduction_(0.0)
| | \ | | | | | - | | | NeighGO_term =
nuclear_organization_and_biogenesis_: signal_transduction_ (0.0)
| | | | | | | | | | | NeighGO_term =
membrane_organization_and_biogenesis_: signal_transduction__ (0.0)
| | | | | | | | | | | NeighGO_term =
organelle_organization_and_biogenesis_: signal_transduction_(0.0)
| | | | | | | | | | | | NeīghGO_term =
nucleolus_: signal_transduction_ (0.0)
| | | | | | | | | | | NeighGO_term =
RNA_binding_: signal_transduction__ (0.0)
| | | | | | | | | | | NeighGO_term =
RNA metabolic process : signal transduction (0.0)
| | | | | | | | | | | NeighGO_term =
ribosome_biogenesis_and_assembly_: signal_transduction_ (0.0)
| | | | | | | | | | | NeighGO_term =
peptidase_activity_: signal_transduction_ (0.0)
| | | | | | | - | | | \ NeighGO term =
microtubule_organizing_center_: signal_transduction_ (0.0)
| | | | | | | | | | | NeighGO_term =
motor_activity_: signal_transduction_ (0.0)
| | | | | | | | | | | NeighGO_term =
cytoskeleton_organization_and_biogenesis_: signal_transduction_
(0.0)
| | | | | | | | | | | NeighGO_term =
cellular_bud_: signal_transduction_ (0.0)
| | | | | | | | | | | NeighGO_term =
site_of_polarized_growth_: signal_transduction_(0.\overline{0})
| | | | | | ` | | | | NeïghGO term =
enzyme_regulator_activity_: signal_transduction_(0.0)
| | | | | | | | | | | NeighGO_term =
cytokinesis_: signal_transduction_ (0.0)
```






```
| | | | | | | | | | start > 95884:
cytoskeleton_organization_and_biogenesis_(14.04)
| | | \ | | |
| | | | | | | | | | | | | | | | | | start <= 111439 
biological_process_ (21.72)
| | | | | - | | | | | start > 110596: sporulation_
(24.74/12.3)
\begin{tabular}{lllllllllll}
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & start \(>111439\) \\
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & start \(<=125128\) \\
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & start \(<=117592\) \\
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\)
\end{tabular}
protein_modification_process_ (16.37)
| | | | | | | | | | | | start > 113765:
ribosome_biogenesis_and_assembly_ (9.68)
| | | | | | | | | | | | | | | | | | | | | | | |tart > 1177592 
organelle_organization_and_biogenesis_ (21.0/12.0)
| | | | | | | | | | | | neigh_num > 1:
vesicle-mediated_transport_(25.58/16.24)
| | | | | | | | | | start > 125128
| | | | | | | | | | | start <= 126831:
biological_process_ (18.09)
| | | | | | | | | | | start > 126831:
protein_modification_process_ (13.94)
| | | | | start > 132427
| | | | | | start <= 151223
| | | | | | strand = C
| | | | | | | start <= 143396
| | | | | | | | | start <= 136691:
cellular_respiration_(13.95)
|
(13.39)
| | | | | | | | | | start > 141250: vesicle-
mediated_transport_ (29.66/11.07)
| | | | | | | | start > 143396
| | | | | | | | start <= 147212:
lipid_metabolic_process_ (19.45)
| |- | | | | - | | | start > 147212:
amino_acid_and_derivative_metabolic_process_ (14.39)
```



```
| | | | | | | | | | | | | | | | | | start <= 143396 
| | | | | | | | | start <= 136001:
biological_process_ (10.75)
| | | | | | | | | | start > 136001:
cell_wall_organization_and_biogenesis_ (13.83)
| | | | | | | | | start > 136691
| | | | | | | | | start <= 141250:
protein_catabolic_process_ (18.6)
| | | | | | | | | | start > 141250:
biological_process_ (7.76)
| | | | | | | | start > 143396
```








```
| | | | | | | | | | | start > 383209:
biological_process_ (17.58)
|
signal_transduction_ (15.78)
| | | | | - | | | | | start > 393118: translation
(23.08)
| | | | | | | | | strand = W:
cell_wall_organization_and_biogenesis_ (28.42)
| | | start > 398271
| | | | start <= 444687
| | | | | start <= 415977
| | | | | | start <= 405245
| | | | | | start <= 400884:
organelle_organization_and_biogenesis_(57.12/34.48)
| | | | | | | - start > 40}00884: conjugation_ (15.14
| | | | | start > 405245
| | | | | | | | strand = C: response_to_stress_
(12.97)
```



```
pseudohyphal_growth_(17.68)
```



```
ribosome_biogenesis_and_assembly_ (20.57)
| | | | | | | | | | start > 411048:
cellular_respiration_(8.96)
| | | | | start > 415977
| | | | | start <= 436015
| | | | | | start <= 426483
| | | | | | | start <= 423035
| | | | | | | | | start <= 419158:
biological_process_ (18.56) | | | | | | start > 419158:
response_to_stress_ (17.74)
| | | | | | | | start > 423035:
DNA_metabolic_process_ (37.39)
| | | | | | | | start > 426483
| | | | | | | start <= 429095
```



```
biological_process_ (8.73)
| | | | | | | | | | start > 427052:
membrane_organization_and_biogenesis_(14.89)
| | \ | | |` |
vitamin_metabolic_process_ (18.53)
| | | | | | | | start > 429095
| | | | | | | | start <= 430945:
response_to_stress_ (40.53/26.75)
```












| | | | | | | 1 start $>728880$ protein_modification_process_ (21.8)
$|\quad| \quad|\quad| \quad \mid \quad$ $\mid$ start > 730382: response_to_stress_
(21.38/9.79)
| | | | | | start > 734634
| | | | | | | start <= 736259: biological_process_ (39.45)
$\begin{array}{llllllll}\mid & \mid & \mid & \mid & \mid & \mid & \text { start }>736259 \\ \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=739836:\end{array}$
organelle_organization_and_biogenesis_ (11.47)
| | | | | | | | | start > 739836:
biological_process_ (22.18)
| | | | | strand $=\mathrm{W}$
| | | | | | start <= 724451
| | | | | | start $<=719028$
| | | | | | | start $<=714450$

biological_process_ (13.25)
| | | | | | | | | start > 704665:
response_to_stress_ (17.84)
| | $\quad$ - | ${ }^{-}$| | | start > 707523:
biological_process_ (17.53)
| | | | | | | start > 714450: meiosis_(23.58)
| | | | | | | start > 719028
| | | | | | | start <= 721385: translation_
(21.31)
| | | | | | | | start > 721385
| | | | | | | | start $<=722606$ :
biological_process_ (21.12)
| | | | | | | | | start > 722606
| | | | | | | | | start <= 723265:
transcription_(21.86)
| | | | ${ }^{-}$| | | | | start > 723265:
biological_process_ (15.85)
| | | | | | start > 724451

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=738577$ |
| :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid \quad$ start $<=728880:$ |  |

ribosome_biogenesis_and_assembly_ (17.08)
| | $\mid$ | | - - | start > 728880
| | | | | | | | start $<=$ 736259:
biological_process_ (29.07)
| | | | | | | | start > 736259:
lipid_metabolic_process_ (29.35/8.07)
| | | | | | - start > 738577
| | | | | | | | distance $<=1329$
| | | | | | | | start <= 739836:
ribosome_biogenesis_and_assembly_ (3.93)
| | | | | | | | $\mid$ start > 739836: translation_ (10.45/0.07)
| | | | | | | distance > 1329:
ribosome_biogenesis_and_assembly_ (30.29)
| start > 744847
| | start <= 779663


## Decision Tree Generated for Chromosome Three:

## For Cellular Component

start $<=265065$
$\mid$
$\mid$
$\mid$








## For Molecular Function



```
| | | | | | | | start <= 16880: lyase_activity_
(14.21)
| | | | | | | | | | | start > 16880: protein_binding_
(37.64)
| | | | | | start > 24325
| | | | | | | start <= 24768: peptidase_activity_
(16.39)
| | | | | | start > 24768: molecular_function_
(16.75)
| | | | start > 27359
| | | | | strand = C: transferase_activity_ (15.2)
| | | | | strand = W
| | | | | | start <= 27929:
transcription_regulator_activity_(33.98/13.61)
| | | | | | -start > \
| | | | | | start <= 31449: transferase_activity_
(8.73)
| | | | | | | start > 31449:
transcription_regulator_activity_ (16.23)
| | | start > 358\overline{65}
| | | | start <= 113768
| | | | | strand = C
| | | | | | start <= 75285
| | | | | | start <= 59026
| | | | | | | | start <= 56527
| | | | | | | start <= 38801:
hydrolase_activity_ (9.31)
| | | | | | | | | start > 38801
| | | | | | | | | | | | | start <= 48364:
molecular_function_ (52.81) | | | | | | | | | start > 48364
| | | | | | | | | | | start <= 52645:
oxidoreductase_activity_ (15.47)
| | | | | | | | | | | start > 52645:
molecular_function_(15.17)
| | | | | | | | start > 56527: RNA_binding_ (18.27)
| | | | | | | start > 59026
| | | | | | | | start <= 65568
```



```
oxidoreductase_activity_(43.53/9.99)
| | | | | | - | | | | distance > 2492:
transferase_activity_ (16.22/5.79)
| | | | | | | | | start > 63441:
molecular_function_(20.94)
| | | | | | | | start > 65568
```



```
| | | | | | | | | start <= 79162:
protein_kinase_activity_ (9.4)
| | | | | | | | | | start > 79162:
molecular_function_(27.09)
| | | | | | start > 90859
| | | | | start <= 107023
| | | | | | | start <= 96281
| | | | | | | | start <= 91324:
oxidoreductase_activity_ (11.91)
| | | | | | | | | start > 91324:
molecular_function_(11.84)
| | |- | | | | | | start > 96281
| | | | | | | | neigh_num <= 1:
transporter_activity_ (10.24/2.04)
| | | | | | | | | neigh_num > 1:
hydrolase_activity_ (20.24/7.55)
| | | | | | | start > 107023
| | | | | | | start <= 109105
| | | | | | | | start <= 108021:
molecular_function_(11.16)
| | | | | | | | | start > 108021:
transferase_activity_ (20.83)
| | | | | | | | start > 109105: molecular_function_
(63.88)
| | | | start > 113768
| | | | | | start <= 122326 
structural_molecule_activity_ (35.24)
| | | | | - | star\overline{t > 118618}
| | | | | | | start <= 120316:
oxidoreductase_activity_ (21.01)
| | | | | | start > 120316: transferase_activity_
(14.32)
| | | | | start > 122326
| | | | | | start <= 127714: molecular_function_ (51.81)
| | | | | | start > 127714
| | | | | | | | | | strand = C 
(16.43)
| | | | | | | | start > 131540:
transporter_activity_ (19.05)
| | | | | | | strand = W: protein_kinase_activity_
(9.92)
| | start > 133122
| | | start <= 162861
| | | | start <= 140931
| | | | | strand = C
| | | | start <= 137744: hydrolase_activity_ (9.77)
nucleotidyltransferase activity (16.4)
| | | | | strand = W: transferase_activity_ (7.92)
| | | | start > 140931
start <= 142544
    | start <= 142168: molecular_function_ (14.78)
```






## For Biological Process

```
| start <= 193293
| | start \(<=50584\)
| | | start \(<=22106\)
| | | | start <= 13282
| | | | | start <= 9706: transport_ (25.02)
| | | | | start > 9706
    | start <= 12285: biological_process (22.61)
    | | start > 12285: transcription_ (30. \(\overline{4} 4 / 5.39)\)
\(\begin{array}{llll}\mid & \mid & \text { start }>13282 \\ \mid & \mid & \mid & \text { strand }=C\end{array}\)
| | | | | | start <= 17290:
amino_acid_and_derivative_metabolic_process_ (18.14)
| \(\left.\left.\left.\right|^{-}\right|^{-}\right|^{-}|\quad| \quad\) s̄art \(>172 \overline{9} 0\) : DNA_Metabolic_process_
(12.34)
| | | | strand = W:
organelle_organization_and_biogenesis_ (14.27)
| | | start > 22106
| | | | start \(<=31449\)
| | | | | start \(<=27359\)
| | | | | start <= 23379:
ribosome_biogenesis_and_assembly_ (22.22)
| | | | | | start > 23379: biological_process_ (84.21)
| | | | | start > 27359
    | start <= 27929: meiosis (14.73)
    | start > 27929:
ribosome_biogenesis_and_assembly_(24.88/10.99)
| | | | start > 31449
| | | | | start \(<=35865\)
| | | | | start \(<=35393\) :
protein_modification_process (11.59)
| | | | | | start > 35393:
cell_wall_organization_and_biogenesis_ (19.48)
| | | | | start > 35865
| | | | \(\quad\) strand \(=C\)
| | | | | | start <= 46905: biological_process_
(45.22)
```










## Decision Tree Generated for Chromosome Four:

## For Cellular Component

| strand = W
| | start <= 142098
| | | start <= 114673
| | | | start <= 36798
| | | | start <= 30657: cellular_component_ (92.25/40.89)
| | | | $\mid$ start > 30657: nucleus_(77.56/36.45)


| | | | | | | | | start > 323471: cellular_component_ (11.72)

| | | | | | | | start > 474043: chromosome
(12.47)
| | | | start > 482264
| | i start $<=603061$
| start <= 558058
| | start $<=497315$
| | | | | | | | | start <= 486801: nucleolus_ (14.64)
| | | | | | | | | start > 486801: ribosome_ (30.44/12.21)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>497315$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | distance $<=3139$ : cytoplasm_ |

(179.84/99.52)
| | | | | | | | | distance > 3139: mitochondrion_ (105.6/68.56)
| | | | | | | start > 558058
| | | | | | | start <= 565925
| | | | | | | | start $<=$ 563525:
plasma_membrane_(26.63)
| | | | | | | | | start > 563525:
endoplasmic reticulum (12.48)
| | | | | | | | start > 565925: cytoplasm_
(56.21/14.77)
| | | | | | start > 570646

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=576471$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=574161$ : mitochondrion_ |

(14.01)
| | | | | | | | start > 574161: membrane_fraction_
(15.44)
| | | | | | | start > 576471
| | | | | | | | start <= 579456: ribosome_ (36.43/21.26)
| | | | | | | | start > 579456
| | | | | | | | start <= 598465
| | | | | | | | $\mid$ start <= 580685:
mitochondrion_ (11.92)













## For Molecular Function


| | | | | | | | | | | | | start > 20635: enzyme_regulator_activity_ (17.39)
| | | | | | | | | | |
protein_binding_(13.8)
$\begin{array}{llllllllll}\mid & - & \mid & - & \mid & \mid & \mid & \mid & \text { start }>28775 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { strand }=\mathrm{W}\end{array}$
molecular_function_ (19.75)
| | | | | ${ }^{-}$| | | | | | | start > 30657:
hydrolase_activity_ (13.28)
| | | | | | | | | |
hydrolase_activity_ (19.06)
$\begin{array}{llllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }>32296 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=33918:\end{array}$
transferase_activity $\quad$ (11.27) $\mid$ | | | | | $\mid$ |
enzyme_regulator_activity_ (14.84)
| | ${ }^{-}$| $\quad$ | | $\left.\right|^{-}$| start > 36478
| | | | | | | | | | strand $=W$
| | | | | | | | | | 1 start<= 36798:
molecular_function_(19.36)
| | | | | | | | | | start > 36798:
protein_binding_ (16.61)
| | | | | | | | | strand = C:
molecular_function_(19.3)
$\left\lvert\, \begin{array}{lllllll}\mid & \mid & \mid & \mid & \mid & \text { start }>42245 \\ \mid & \mid & \mid & \mid & \mid \quad \text { start }<=42701 \text { : }\end{array}\right.$
phosphoprotein_phosphatase_activity_ (13.03)
| | | | | | | | | - start > 42701
| | | | | | | | | | start <= 48032
| | | | | | | | | | 1 start<=44066:
protein_binding_(11.35)
| | | | | | | | | | | start > 44066:
hydrolase_activity_ (17.27)
| | | | | | | | | start > 48032:
protein_binding_ (16.19)
| | | | | | start > 52174
| | | | | | start $<=77967$
| | | | | | start $<=66494$
| | | | | | | start $<=56347$
| | | | | | | | | start <= 52446:
structural_molecule_activity_ (12.42)
$\mid$ | | | | | | | | | start > 52446: RNA_binding_ (14.53)
$\begin{array}{llllllll}\mid & \mid & \mid & \mid & \mid & \mid & \text { start }>56347 \\ \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=61802 \text { : }\end{array}$
molecular_function_(35.36)
| | | | | | | | | | start > 61802

hydrolase_activity_ (10.24)
| | | | | | | | | | ${ }^{-}$| start > 65243:
molecular_function_(6.77)









```
| | | | | | | | start <= 394214:
helicase_activity_ (13.8)
| | | | | | | | | start > 394214: RNA_binding_
(12.97)
| | | | | | | | start > 397534:
structural_molecule_activity_ (17.37)
| | | | | - | | start > 403288
| | | | | | | | start <= 416705
| | | | | | | | | start <= 411822
| | | | | | | | | | start <= 410376:
| | | | | | - | | | | | start > 410376:
oxidoreductase_activity_ (5.52)
| | | | | | | | | start > 411822:
molecular_function_(16.37)
| | | | | | | | start > 416705
| | | | | | | | | start <= 424206:
protein kinase activity (19.1)
| | | | | | | | | start > 424206:
transferase_activity_ (15.35)
| | | | | | | | start > 427361 
| | | | | | | start <= 439906
| | | | | | | | | start <= 437769
| | | | | | | | | | start <= 429064:
protein_binding_ (17.15)
```



```
| | | | | | |art <= 432327:
molecular_function (13.4)
| | | | | | | | | | | | start > 432327:
protein_binding_ (11.63)
| | | | | | | | | | | start > 434333:
molecular function (15.45)
| | | | | | | | | start > 437769
| | | | | | | | | | start <= 438044:
peptidase_activity_ (11.22)
| | | | | | | | | | start > 438044:
phosphoprotein phosphatase activity (22.39)
| | | | | | | | | star̄t > 439906
| | | | | | | | start <= 443026:
transporter_activity_ (16.4)
| | | | | | | | | start > 443026
| | | | | | | | | start <= 454779
| | | | | | | | | | start <= 453042
| | | | | | | | | | | | start <= 447981:
molecular_function_(23.34)
| | | | | | | | | | | | start > 447981:
protein_binding_ (15.46)
| | | | - | | | | | | start > 453042:
molecular function_(32.24)
| | | | | | | | | | start > 454779:
protein_binding_ (11.81)
| | | | | | | start > 458100
```









```
| | | | | | | start > 767968: isomerase_activity_
(19.83)
| |lllllol
protein_binding_ (23.2)
| | | | | | | | | | start > 771874:
molecular_function_(51.8)
| llllllllllll
| | | | | | | | | | | start<= 784868:
protein_binding_ (17.02)
| | | | | | | | | | | start > 784868:
transcription_regulator_activity_ (15.64)
| | | | | | - | | \ | start > 789446:
protein_binding_ (30.65)
| | | | | | | start > 806618
| | | | | | | | start <= 808321:
translation regulator activity (24.6)
| | | | | | | | start > 808321:
transcription_regulator_activity_ (46.25)
| | | | | | | start > 814449
| | | | | | | start <= 829148
| | | | | | | | start <= 817947
| | | | | | | | | | start <= 816875:
molecular_function_(14.76)
| | | | | | | | | | start > 816875:
oxidoreductase_activity_ (21.27)
| | | | | | | | | start > 817947:
molecular_function_(70.68)
| | | | | | | | start > 829148: protein_binding_
(43.51)
| | | | | | strand = C
| | | | | | start <= 802219
| | | | | | | start <= 794720
| | | | | | | | start <= 790325
| | | | | | | | | start <= 781420:
molecular_function_(11.98)
| | | | | | | | | | start > 781420:
protein_binding_ (29.71)
| | | | | | | | | start > 790325:
molecular_function_(22.17)
| | | | | | | | start > 794720:
enzyme_regulator_activity_ (18.77)
| | | | | | | start > 802219
| | | | | | | start <= 838389
| | | | | | | | start <= 829148
| | | | | | | | | | start <= 806618:
molecular_function_(18.42)
| | | | | | | | | | start > 806618
```














$\mid$ | | | | | | | | | | | | | strand $=W$ : molecular_function_(29.43)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ strand $=C$ : protein_kinase_activity_ (17.87)
 1436922: structural molecule_activity_ (22.15) $|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ | $\mid$ | $\mid$ | $\mid$ 1436922: molecular_function_(21.6)
| | | | | | | | | | | start > 1441142: RNA_binding_(21.76)
 (53.48)


| | | | | | | | start > 1523234: helicase_activity_ (4.51)

## For Biological Process

```
| start <= 241418
| | start <= 130485
| | | | start <= 78427
| | | | | start }<=3065
| | | | | start <= 16204: transport_(68.13/44.38)
| | | | | | start > 16204: biological_process_
(124.9/48.54)
| | | | | start > 30657
| | | | | start <= 76546
| | | | | | | strand = W
| | | | | | | | start <= 42701:
protein_modification_process_ (69.89/22.8)
| | | | | \ | ` start > 42701: translation_
(58.8/31.43)
| | | | | | | strand = C
| | | | | | | | start <= 52446: vesicle-
mediated_transport_(134.81/69.44)
| | | | | | | | start > 52446
| | | | | | | start <= 68607
| | | | | | | | | start <= 60406:
biological_process_(43.16/28.39)
| | | | | | | | | | start > 60406:
membrane_organization_and_biogenesis_(40.57/24.23)
| | | | | | | | | start > 68607
| | | | | | | | | | start <= 70320:
protein_modification_process_ (37.58/16.5)
| | | | | | | | | | start > 70320:
biological_process_ (39.94/17.45)
| | | | | start > 76546: transport_ (56.06/34.63)
| | | | start > 78427
| | | | start <= 88249
| | | | | | strand = W: transport_ (94.03/57.52)
| | | | | | strand = C: RNA_metabolic_process_
(35.85/16.11)
| | | | start > 88249: biological_process_ (68.27)
| | | start > 97954
| | | | start <= 116322
| | | | | start <= 101068
| | | start <= 98476: translation_(20.95)
(25.57/11.04)
| | | | | start > 101068
| | | | | strand =W
| | | | | | | start <= 114673
| | | | | | | | start <= 107209: vesicle-
mediated transport (23.7/11.25)
| | | | | | | | start > 107209: signal_transduction_
(43.14/16.42)
```





```
| | | | | | | | start <= 373605:
biological_process_ (40.0/15.23)
| | | | | | | | start > 373605: translation_
(59.12/40.43)
| | | | | | | start > 382330
biological_process_ (22.23)
| | | | | | | | | start > 385584:
RNA_metabolic_process_ (9.9)
| | | start > 388898
| | | | start <= 709546
| | | | | start <= 569768
| | | | | start <= 461244
| | | | | | | strand = W
| | | | | | | | start <= 432327
| | | | | | | | | | | | | | start <= 42 | | | | | | | | | | | | | | | | | 
RNA_metabolic_process_(53.5/40.16)
| | | | | | | | | | start > 411822:
biological_process_ (56.95/31.96)
| | | | | | | | | start > 427361:
response to_stress_ (38.34/17.69)
| | | | | - | | | start > 432327
| | | | | | | | start <= 438044
| | | | | | | | | | start <= 437769: cell_cycle_
(28.56/13.47)
| | | | | | | | | | start > 437769:
protein_catabolic_process_ (21.13)
| | - | | | | | | | start > 438044
| | | | | | | | | start <= 443026
| | | | | | | | | | | start <= 439906:
organelle_organization_and_biogenesis_ (48.34/28.36)
| | | | | | | | | | | start > 439906:
generation_of_precursor_metabolites_and_energy (21.16)
| | | - | | | - | | | - | - start > 443026
| | | | | | | | | | | start <= 453042:
cell_cycle_(40.86/12.03)
| | | | | | | | | | | start > 453042:
biological_process_(41.12/24.46)
| | | | | | | strand = C
| | | | | | | start <= 416705
| | | | | | | | | | | | | | | | inart <= 411822 
(36.25/23.09)
| | | | | | | | | | start > 403288:
biological_process_ (29.38/8.42)
| | | | | | | | | start > 411822:
DNA_metabolic_process_(55.0/38.39)
| | | | | | | | | start > 416705
| | start <= 421511: vesicle-
mediated_transport_ (16.87)
| | | | | | | | | | start > 421511:
lipid_metabolic_process_(21.16/9.55)
```




| | | | | | | | | | | start > 668077: biological_process_ (40.74)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>679761$ |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=681614:$ |
| DNA metabolic process | $(34.87 / 10.43)$ |  |  |  |  |  |  |  |  |  |

DNA_metabolic_process_(34.87/10.43) | | | | | | | | | | | |
biological_process_(42.78/10.64)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>688224$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\|\quad\| \quad$ distance $<=1628:$ |  |  |

ribosome_biogenesis_and_assembly_ (15.31/5.47)
$||||||||\mid$ distance $>$ 1628:
cell_wall_organization_and_biogenesis_ (54.5/34.32)
| $\quad$ |- | start $>70 \overline{9} 546$
| | | | | start <= 956009
$|\quad| \quad \mid \quad$ start $<=930353$
| | | | | | start $<=858134$
| | | | | | | | start $<=784868$
$\begin{array}{lllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=776160 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=746735:\end{array}$
DNA metabolic process (113.73/79.36)

| $\mid$ | - | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>746735$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=769522$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=765703$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=$ |
| $751628: ~ l i p i d ~ m e t a b o l i c ~ p r o c e s s ~$ | $(38.23 / 23.87)$ |  |  |  |  |  |  |  |  |  |

751628: lipid_metabolic_process_(38.23/23.87) | | | | | | | | | | | | | |
751628: cytoskeleton_organization_and_biogenesis_(47.31/29.38)
$\mid$ | | | $\mid$ | | | ${ }^{-}$| | $\mid$start > 765703:
transcription_(27.99/11.78)
$\begin{array}{llllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }>769522 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\ \text { start }<=771874:\end{array}$
cytoskeleton_organization_and_biogenesis_(54.03/28.41)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ start $>$ 771874:
biological_process_ (18.3)
$\begin{array}{llllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }>776160 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=779040:\end{array}$
ribosome_biogenesis_and_assembly_ (19.6)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ start $\mid$ 779040:
RNA_metabolic_process_(29.83/13.09)

| $\mid$ | - | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>784868$ |  |
| ---: | ---: | ---: | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=808321$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| \| | start $<=794720:$ |  |  |  |  |  |  |  |  |

cell_cycle_(113.83/74.03)
| $\mid$ | | | | | | | | start > 794720:
translation_(55.24/28.85)
| | | | | | | | | | start > 808321

| | | | | | | | | | | | neigh_num <= 1:
biological_process_(49.16/28.29)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ neigh_num > 1:
DNA_metabolic_process_(83.93/52.82)





```
| | | | | | | | | | | start > 1080195:
DNA_metabolic_process_ (106.5/54.14)
| | | | | | | | | | start > 1092508:
biological_process_ (77.19/35.14)
| | | | | | start > 1105824
| | | | | | | | | start <= 1117121
| | | | | | | | | | | start <= 1112290:
translation_(40.53/19.71)
| | | | | | | | | | | start > 1112290:
cell_cycle_(13.53)
| | | | | | | | | | start > 1117121:
protein_catabolic_process_ (18.22)
| | - | | | |- | | strand = C
| | | | | | | | | | start <= 1114024:
cell_cycle_(57.41/33.72)
| | | | | | | | | | start > 1114024
| | | | | | | | | | distance <= 3583:
cytokinesis_(43.51/28.83)
| | | | | | | | | | | distance > 3583:
biological_process_(23.43/11.92)
| | start > 1127867
| | | strand = W
| | | | distance <= 7637
| | | | | start <= 1414567
| | | | | | | start <= 1183292 
protein modification process (20.28)
| | - | | \ | start > 1129583
| | | | | | | | start <= 1175823
| | | | | | | | | | | | | | | | | distance <= 3358:
biological_process_ (28.79/1.38)
| | | | | | | | | | | distance > 3358:
DNA_metabolic_process_(26.61/13.35)
| | | | | | | | | start > 1145086: translation_
(27.66)
| | | | | | | | start > 1175823
| | | | | | | | start <= 1178659:
cell_wall_organization_and_biogenesis_(14.37)
| | | | | | | | | start > 1178659:
response_to_stress_(21.1/9.55)
| | | | | | start > 1183292
| | | | | | start <= 1259893
| | | | | | | start <= 1228603
| | | | | | | | start <= 1206375: vesicle-
mediated_transport_(83.86/57.06)
| | | | | | | | | start > 1206375:
biological_process_(107.91/56.57)
| | | | | | | | | start > 1228603
| | | | | | | | start <= 1246076
| | | | | | | | | | start <= 1234210:
pseudohyphal_growth_(43.07/17.5)
| | | | | | | | | | start > 1234210
```



| | | | | | | | | | start > 1149460: organelle_organization_and_biogenesis_ (26.22/11.89)
| | | | | $\left.\left.\right|^{-}\right|^{-}|\quad| \quad$ start > 1150514: translation_ (27.75)

$|$| $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>1153620$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=1169172$ : transport_ | (45.75/20.33)

| | | | | | | | start > 1169172:
biological_process_(52.05)
| | | | | | | start > 1190052
| | | | | | | start <= 1194877: DNA_metabolic_process_(9.66)
$\begin{array}{lllllllll}- & \mid & \mid & \mid & \mid & \mid & \text { start }>1194877 \\ \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=1199175 \text { : }\end{array}$
ribosome_biogenesis_and_assembly_ (38.03/15.22)
$|\quad| \quad|\quad| \quad$ | $|\quad| \quad$ | start > 1199175: cell_cycle_ (38.24/14.31)

$|$| $\mid$ | $\mid$ | $\mid$ |
| :--- | :--- | :--- |
| $\mid$ | $\mid$ | start $>1206375$ |
| $\mid$ | $\mid$ | start $<=1219405$ : biological_process_ |

(56.33/16.68)

biological_process_ (14.88) | | | | | | | |
cellular respiration (18.94/8.69)
$|\quad| \quad|\quad| \quad \mid \quad$ | $\mid \quad$ start $>$ 1226814:
biological_process_(83.95/34.43)


```
translation (17.49)
```

| | | | | | | | | | start > 1257071

| | | | | | | | | | start > 1261673:
transcription_(94.59/61.65)
| | | | | | | | start > 1289398:
protein_modification_process_(51.36/18.28)
| | $|\quad| \quad|\quad| \quad$ | start > 1296678
| | | | | | | start $<=1319267$
| | | | | | | | | start <= 1306259:
biological_process_ (21.95)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ start > 1306259:
signal_transduction_(25.25/8.0)


```
| | | | | | | start <= 1483395: sporulation_
(84.43/43.11)
| | | | | | | | start > 1483395:
cell_wall_organization_and_biogenesis_ (102.91/52.88)
| | | | | | | start > 14}87030: transport_
(31.85/13.41)
| | | | | start > 1488982
| | | | | start <= 1496540
| | | | | | | start <= 1494578: RNA_metabolic_process_
(35.77/18.53)
| | | | | | start > 1494578: electron_transport_
(39.38/19.65)
| | | | | | start > 1496540: biological_process_
(87.06/43.89)
```


## Decision Tree Generated for Chromosome Five:

## For Cellular Component

|
start $<=166884$
$\mid$
$\mid$







```
| | | | | | | | NeighGO_term =
nucleotidyltransferase_activity_: membrāne_ (0.0)
| | | | | | | | NeighGO_term = transcription_:
membrane_ (0.0) | | | | | NeighGO_term = response_to_stress_:
membrane (2.55/1.05)
| | \ | | | | | NeighGO_term =
structural_molecule_activity_: membrane_(0.0)
| | | | | - | | \ NeighGO_term = translation_:
membrane_(0.0)
| | | | | | | | NeighGO_term = RNA_binding_:
membrane_(0.0)
| | \ | | | | | NeighGO_term =
RNA_metabolic_process_: membrane_ (0.0)
| | | | | | | | NeighGO_term =
ribosome_biogenesis_and_assembly_: membrane_ (0.0)
| | | | | | | | NeighGO term =
oxidoreductase_activity_: endoplasmic_rèticulum_(1.95/1.01)
| | | | | | | | Neigh\overline{GO_term = \overline{electron_transport_:}}\mathbf{l}=\mp@code{l}
membrane_ \ (0.0) | | | | NeighGO_term = DNA_binding_:
membrane (0.0)
| | \ | | | | | NeighGO_term =
protein_catabolic_process_: membrane_(\overline{0}.0)
| | | | | | | | NeighGO_term = ribosome_: membrane_
(0.0)
| | | | | | | | NeighGO_term =
cellular_respiration_: membrane_ (0.0)
| | | | \ | | -NeighGO term =
enzyme_regulator_activity_: membrane_(\overline{0.0)}
| | | | | | | | NeighGO_term =
cytoskeleton_organization_and_biogenesis_: membrane_ (0.0)
| | | | | | | | NeighGO_term = vesicle-
mediated_transport_: membrane_ (0.0)
| | \ | | | | | N NeighGO_term = hydrolase_activity_:
endoplasmic_reticulum_(2.19/0.83)
| | | | | | | | NeighGO_term =
carbohydrate_metabolic_process_: membrane_(0.0)
| | | \ | | | | | - NeighGO_tērm =
amino_acid_and_derivative_metabolic_process_: membrane_ (9.29/5.11)
| | | | | | | | NeighGO_term =
cell_wall_organization_and_biogenesis_: endomembrane_system_
(2.64/0.42)
| | | | | | | NeighGO_term = isomerase_activity_:
membrane_(0.0)
| | | | | | | | NeighGO_term = endomembrane_system_:
membrane_ (0.0) | | | | | NeighGO_term = cytokinesis_:
membrane_(0.0)
| | \ | | | | | NeighGO_term =
microtubule_organizing_center_: membrane_ (0.0)
| | | | | | | | NeighGO_term = peptidase_activity_:
membrane_(0.0)
```











For Molecular Function











```
| | | | | | start <= 396765: molecular_function_
(85.46)
| |llllllll
signal transducer activity_ (11.14)
| | | | |- | | | | | start > 399051:
transporter_activity_ (11.38)
| | | | | \ | | start > 402371:
enzyme_regulator_activity_ (14.95)
| | | | | | start > 404809
| | | | | | start <= 430445: molecular_function_
(71.87)
| | | | | | | start > 430445
| | | | | | | | start <= 432491: RNA_binding_
(18.59)
| | | | | | | | start > 432491: molecular_function_
(63.8)
| [|:llllll
(15.49)
| | | | | | | | start > 390048:
structural_molecule_activity_ (8.87)
| | | | | | | | start > 396765: protein_binding_
(20.79)
```



```
molecular_function_(15.71)
| | | | | | | | | | | start > 402371:
protein_kinase_activity_ (20.29)
| | | | | | | | | | start > 404809
| | | | | | | | | | start <= 414477
| | | | | | | | | | start <= 410185:
molecular_function_ (18.56)
| | | | | | | | | | | | start > 410185:
RNA_binding_(15.49)
| | | | | | | | | | | start > 414477:
molecular_function_(19.52)
| | | | | | | | | start > 415855:
protein_kinase_activity_ (12.44)
| | | | | | | | start > 417277
| | | | | | | | start <= 423948:
structural_molecule_activity_ (14.29)
| | | | | - | | ` | start > 423948: RNA_binding_
(21.5)
| | | | | | start > 424694
| | | | | | | start <= 434608:
phosphoprotein_phosphatase_activity_ (15.13)
```





## For Biological Process

start $<=364585$
$\mid$
$\mid$
$\mid$


| | | | | | | | | | | start <= 120299: biological_process_(11.42)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ start > 120299:
DNA_metabolic_process_(28.26)
| | | | | | | | | | | | start <= 122929: transport (10.09)
| | | | | | | | | | | | | start > 122929:
biological_process_ (36.44)
| | | | | | | | | | start > 126218
| | | | | | | | | | | start <= 126629:
RNA_metabolic_process_ (16.9)
$\begin{array}{llllllllll}- & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }>126629 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid \quad \text { neigh_num }<=1:\end{array}$ organelle_organization_and_biogenesis_ (15.72/5.42)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ neigh_num $>1$ | | | distance <= 2351: membrane organization and biogenesis (5.16/1.21)
| | | | | | | | | | | $\left.\right|^{-}|\quad|$ distance > 2351:
organelle_organization_and_biogenesis_(6.88/2.06)

carbohydrate_metabolic_process_(8.41)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ start > 131772:
generation_of_precursor_metabolites_and_energy_ (26.49/11.14)

(10.01)
| | | | | | start > 138666

| | $\left.{ }^{-}\right|^{-}|\quad| \quad|\quad| \quad \mid \quad$ start $>^{-} 141891$
$\begin{array}{llllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=150013 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \\ \text { start }<=148175:\end{array}$
biological_process_(12.8)
| | | | | | | | | | | start > 148175:
cell_cycle_(11.91)
| | | | | | | | | start > 150013:
biological_process_ (18.84)
| | | | | | | | strand = W
| | | | | | | | start $<=$ 141891:
biological_process_ (16.88)
| | | | | | | | | start > 141891
$\begin{array}{llllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=146950 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=144326 \text { : }\end{array}$
transport (16.23)
| | | | | | | | | | | start > 144326:
cell_wall_organization_and_biogenesis_ (40.37/21.92)
| | | | | | | | | start > 146950:
biological_process_(14.42)





```
| | | | | | | | | start <= 292064:
biological_process_ (33.5)
| | | | | | | | | | start > 292064
| | | | | | | | | | start<= 293048:
conjugation_(9.43)
| | | | | | | | | | | start > 293048:
transcription (14.11)
| | | | | | | | start > 295408:
DNA_metabolic_process_ (24.16)
```



```
biological_process_ (9.91)
```



```
transport (45.45/29.01)
| | | | | | | | | | | start > 302804:
electron_transport_ (12.35)
|
protein_modification_process (13.46)
| | | | | | | ` | | | start > 311195:
translation_(8.56)
```



```
translation (22.95)
| | | | | | | | | | | start > 306319:
transport (21.47/8.48)
| | | | | | | | | | | | | | | | start <= 319959:
biological_process_ (80.72)
| | | | | | | | | start > 319959: transcription
(21.23)
| | | | | start > 322208
| | | | | | start <= 348396
| | | | | | | start <= 327027
```



```
ribosome_biogenesis_and_assembly_ (13.58)
| | | | | | | | | start > 325932:
cell_wall_organization_and_biogenesis_ (18.96)
| | | | | | | | strand = W:
amino_acid_and_derivative_metabolic_process_ (13.99)
```



```
transport_ (17.14)
```

$\mid$ | | | | | | | $\mid$ start > 332826: biological_process_ (14.68)

amino_acid_and_derivative_metabolic_process_ (16.19)
| $\left.\left.\left.\left.\right|^{-}\right|^{-}\right|^{-}|\quad|\right|^{-}|\quad| \quad \mid \quad{ }^{-}$start > 328473: biological_process (7.94)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>335184$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid \quad$ start $<=337336:$ |

response_to_stress_ (9.37)
| | | | | | | | | | start > 337336:
amino_acid_and_derivative_metabolic_process (19.38)

| $\mid$ | $\mid$ | $\mid$ | $\left.\right\|^{-}$ | $\mid$ | $\mid$ | $\mid$ | start $>342 \overline{1} 63$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| $\mid$ | $\mid \quad$ strand $=C$ |  |  |  |  |  |  |
| itart $<=347608:$ |  |  |  |  |  |  |  |

cell_wall_organization_and_biogenesis_ (12.36)
| $\left.\left|\left.\right|^{-}\right| \quad\left|\left.\right|^{-}\right|^{-}|\quad|\right|^{-}$start > 347608:
biological_process_(11.3)
| | | | | | | | | strand = W: biological_process_ (26.05)
| | | | | | start > 348396
| | | | | start $<=359558$

| 1 | $\mid$ | $\mid$ | $\mid$ | start $<=354930$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | strand $=C:$ |

protein_catabolic_process_(8.0)
| | | | | | | | | strand = W
| | | | | | | | | start <= 349976:
DNA_metabolic_process (13.2)
| ${ }^{-}$| | ${ }^{-}$| | | | ${ }^{-}$start > 349976 | | | | | | | | | | | start $<=$ 351694: sporulation_(14.04)

translation (4.69)
| | | - | | | | start > 354930
| | | | | | | | strand = C:
amino_acid_and_derivative_metabolic_process_ (12.69)
| | | | | | | | strand = W:
protein_modification_process_(39.14/10.93)
| | - | | $\quad \mid \quad \overline{s t a r t ~>~} 359558$
| | | | | | | | strand = C: biological_process_
(25.61)
| | | | | | | strand $=\mathrm{W}$
| | | | | | | start <= 363096: translation_
(16.85)
| | | | | | | | start > 363096:
response_to_stress_ (16.41)
| start > 364585
| | start <= 414477
| | | start <= 374541
| | | | strand = C
| | | | | start <= 372322:
nuclear_organization_and_biogenesis_ (26.19)
| | | | | start > 372322:
membrane_organization_and_biogenesis_ (40.17/18.7)






## Decision Tree Generated for Chromosome Six:

## For Cellular Component








## For Molecular Function

| start $<=153119$
$\mid$
$\mid$
$\mid$
$\mid$
$\mid$
$\mid$$|$






## For Biological Process

| start <= 55985
| | start <= 15431
| | | start <= 9545: biological_process_ (88.17)







## Decision Tree Generated for Chromosome Seven:

## For Cellular Component

start $<=299731$
$\mid$
$\mid$
$\mid$








```
| | | | | | | | NeighGO_term = peroxisome_:
chromosome (0.0)
| | | | | | | | NeighGO_term =
lipid_metabolic_process_: chromosome_(\overline{0.0)}
| | | | | | | | NeighGO_term = DNA_binding_:
chromosome (0.0)
| | | | | | | | NeighGO_term = helicase_activity_:
chromosome (0.0)
| | | | | | | | NeighGO_term =
DNA_metabolic_process_: nucleus_ (1.34/0.67)
| | | | | | | | _-NeighGO_term = meiosis_: chromosome_
(0.0)
| | | | | | | | NeighGO_term =
phosphoprotein_phosphatase_activity_: chromosome_ (0.0)
| | | | | | | | NeighGO_term = pseudohyphal_growth_:
chromosome_ (0.0)
| | | | | | | | NeighGO_term = transcription_:
chromosome (0.0)
| | | | | | | | NeighGO_term =
protein_kinase_activity_: chromosome_ (0.0)
| | | | | | | | NeighGO_term = RNA_binding_:
chromosome_ (0.0)
| | | | | | | | NeighGO term =
cell_wall_organization_and_biogenesis_: chromosome_ (0.0)
| | | | | | | | NeighGO_term = cytokinesis_:
chromosome_ (0.0)
| | | | | | | | NeighGO_term = endomembrane_system_:
chromosome_ (0.0)
| | | | | | | | NeighGO_term =
nuclear_organization_and_biogenesis_: chromosome_ (0.0)
| | | | | | | | NeighGO_term =
microtubule_organizing_center_: chromosome_ (0.0)
| | | | | | | | NeighGO_term = sporulation_:
chromosome (0.0)
| | | | | | | | NeighGO_term =
cellular_respiration_: chromosome_ (0.0)
| | | | | | | | NeighGO_term = ribosome_:
chromosome (0.0)
| | | | | | | | NeighGO_term = lyase_activity_:
chromosome (0.0)
| | | | | | | | NeighGO_term =
amino_acid_and_derivative_metabolic_process_: chromosome_ (0.0)
| | | | | | | | NeighGO_term =
nucleotidyltransferase_activity_: chromosome_(0.0)
| | | | | | | NeighGO_term }\mp@subsup{}{}{-}=\mathrm{ nucleolus_: nucleus_
(4.4/1.5)
| | | | | | | NeighGO_term = cellular_bud_:
chromosome_(0.0)
| | | | | | | | NeighGO_term =
site_of_polarized_growth_: chromosome_ (0.0)
| | | | | | | | NeighGO_term =
extracellular_region_: chromosome_(0.0)
| | | | | | | | NeighGO_term = conjugation_:
chromosome_(0.0)
```










| | | | | | | | | | | | | | start $>$ 466146: nucleus_(16.8)
| | | | $\mid$ | | | | | | | 1 start $>472860$ | | | | | | | | | | | | | | start < = 475550: mitochondrion_(16.45)
| | | | | | | | | | | | | | start > 475550 | | 485925
 start $<=482671$
| | | | | | | $\mid$ | $\mid$ | $\mid$ | | | start <= 478657: cellular_component_ (12.59)
| | | | | | | | | 1 | | | | | | start > 478657: plasma_membrane_(39.71/23.91)

| | | | | $\mid$ | | $\mid$ | | | | |

485925: mitochondrion_(10.69)
$\begin{array}{lllllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }>490551 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=494521 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\ \text { start }<=491960:\end{array}$ cytoplasm (18.02)
| | | | | | | | | | | | start > 491960:
nucleus_(46.71/17.95)
| | | | | | | | | | | start > 494521:
endoplasmic_reticulum_ (20.78)

(15.0)
| | | | | | | | | start > 498038
| | | | | | | | | | | start $<=499910$ : nucleus_(14.46)
| | | | | | | | | | | start > 499910: cytoplasm_(33.89/15.46)
| | | | | | | | start > 500136: peroxisome_ (16.85)
| | | | | | | | start > 502942
| | | | | | | | strand = W: nucleus_
(58.23/15.32)
| | | | | | | | strand = C: cellular_component_
(30.36/13.67)
| | | | | | start > 506973
| | | | | | | start <= 522080
| | | | | | | | strand = W

| | | | | | | | | | | start $<=$ 511549:
nucleus_(18.57)
| | | | | | | | | | start > 511549:
mitochondrion_(12.25)







| | | | | | | | | | | | | | neigh_num > 1: chromosome_(22.18/8.01)
| | | | | | | | | | | | | | start > 767434: cellular_component_ (16.33)

| | | | | | | | | | | | | | start >
774354
| | | | | | | | | | | | | | | start <=
780404

| | | | | | | | | | | | | | | | |
start <= 775198: membrane_(11.12)
| | | | | | ${ }^{-}$| |
start > 775198: cytoplasm_(11.55)
| | | | | | | | |
start > 779621: mitochondrion_ (13.07)
| | | | | | | | | |
780404
| | | | | | | | | | | | | | |
start $<=784863:$ nucleolus_(12.06)
| | | | | | | | | | | | | | | |
start > 784863
| | | | | | | | |
start $<=$ 789036: membrane_ (15.77)
| | | | | | | ${ }^{-}$|
start > 789036: cytoplasm (11.92)
| | | | | | | | | | | | start > 794422:
cellular_component_ (17.7)
| | | | | | | | | | strand = C
$\mid$ | | | | | | | | | | $\left\lvert\, \begin{aligned} & \text { strand }=C \\ & \text { start }<=786925\end{aligned}\right.$
| | | | | | | | | | | | | start <= 750405:
nucleolus_(15.24)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ start $>750405$
| | | | | | | | | | | | | start <=
775198

| | | | | | | | | | | | | | | |
start $<=754731$ : cytoplasm_(11.55)
| | | | | | | ${ }^{-}$| | | | | | | |
start > 754731: mitochondrion_(26.05)
| | | | | | | | | | | | | | | start >
757775: plasma_membrane_(12.53)
| | | | | | | | | | | | | | start >
775198
| | | | | | | | | | | | | | start <=
785442: cellular component_(26.82)

785442

| | | | | | | | | | | | | | | NeighGo_term = transcription_regulator_activity_: mitochondrion_ (0.0)
| | | | | | | | | | | | | | |
NeighGo_term = signal_transduction_: mitochondrion_ (0.0) | | | | | | | | | ${ }^{-}$| | | | NeighGO_term $=$ nucleus_: nucleus (8.39/3.69)
| | | | | | | | | | | | | | | NeighGO_term = enzyme_regulator_activity_: mitochondrion_ (0.0)
 NeighGO_term = ribosome_biogenesis_and_assembly_: nucleolus_ (5.38/2.87)
| | | | | | | | | | | | | | |
NeighGO_term = ligase_activity_: mitochondrion_ (0.0)
| | | | | | | | | | | | | | | |
NeighGO_term = translation_: mitochondrion_ (0.0)
| | | | | | | | | | | | | | |
NeighGO_term = protein_binding_: mitochondrion_(0.0)
| | | | | | | | | | | | | |
NeighGO_term = transport_: mitochondrion_(0.0)
NeighGo_term = endoplasmic_reticulum_: mitochondrion_ (0.0)
$|\quad| \quad$ | $|\quad| \quad|\quad| \begin{array}{lllllll}- & \mid & \mid & \mid & \mid & \mid & \mid\end{array}$
NeighGO_term = membrane_: mitochondrion_ (0.0)
| | | | | | | | | | | | | | | |
NeighGO_term = Golgi_apparatus_: mitochondrion_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad|$
NeighGO_term = transporter_activity_: mitochondrion_(0.0)
| | - | | | | ${ }^{-}$| |
NeighGo_term = hydrolase_activity_: mitochondrion_(4.77/1.36)

mitochondrion_(0.0)
| | | | ${ }^{-}$| | | | | | | | | |
NeighGO_term = motor_activity_: mitochondrion_ (0.0)

NeighGO_term = cell_cycle_: mitochondrion_ (0.0)
NeighGO_term = cytoskeleton_organization_and_biogenesis_:
mitochondrion_(0.0)
| | | | | | | | | | | | | | | |
NeighGO_term = chromosome_: mitochondrion_ (0.0)
NeighGo_term = RNA_metabolic_process_: mitochondrion_(0.0)

NeighGO_term = vacuole_: mitochondrion_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid$
NeighGO_term = vesicle-mediated_transport_: mitochondrion_(0.0)

NeighGo_term = cytoplasmic_membrane-bound_vesicle_: mitochondrion_
(0.0)

NeighGO_term = structural_molecule_activity_: mitochondrion_(0.0)


$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid$
NeighGO_term = extracellular_region_: mitochondrion_ (0.0)

$|\quad|{ }^{-}|\quad| \quad|\quad|{ }^{-}|\quad| \quad| |^{-}|\quad| \quad|\quad| \quad \mid$
NeighGO_term = response_to_stress_: mitochondrion_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid$
NeighGO_term = plasma_membrane_: mitochondrion_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid$
NeighGO_term = isomerase_activity_: mitochondrion_ (0.0)

NeighGO_term = cell_budding_: mitochondrion_ (0.0)
| | | | | | | | | | | | | | | |
NeighGo_term = translation_regulator_activity_: mitochondrion_(0.0)
| | | | | | | | | | | | | | | |
NeighGo_term = membrane_fraction_: mitochondrion_(0.0)
| | | | | | $\mid$ | $\mid$ | | | 1
NeighGO_term = cell_wall_: mitochondrion_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ | | | | | |
NeighGO_term = peptidase_activity_: mitochondrion_ (0.0)
| | - | | | $\quad$ | $\left\lvert\, \begin{aligned} & - \\ & \mid \\ & \mid\end{aligned}\right.$
NeighGO_term = protein_catabolic_process_: mitochondrion_ (0.0)
| | | | | | | | | | | | | | | |
NeighGo_term = mitochondrial_envelope_: mitochondrion_(0.0)
NeighGO_term = vitamin_metabolic_process_: mitochondrion_(0.0)
$|\quad| \quad|\quad| \quad|-|\quad| \quad| \quad|\quad| \quad|\quad|$
NeighGO_term = signal_transducer_activity_: mitochondrion_(0.0)
| | | | | | | | | | | | | |
NeighGO_term = membrane_organization_and_biogenesis_: mitochondrion_
(0.0)
| | | | | | | | | | | | | | |
NeighGO_term = carbohydrate_metabolic_process_: mitochondrion_(0.0)
| | | | | | | | | | | | | | | |
NeighGO_term = anatomical_structure_morphogenesis_: mitochondrion_
(0.0)
| | | | | | | | | | | | | | |
NeighGo_term = cell_cortex_: mitochondrion_ (0.0)

| | | | | | | | | | | | | | | start >
807661: cytoplasm_(18.02/5.77)
| | | | | | | | | | | start > 809425
| | | | | | | | | | $\mid$ start $<=829121$
$||||||||||\mid$ start $|=$ 810232:
cellular_component_(8.26)
| | | | | | | | | | | | | start > 810232
| | | | | | | | | | | | | start < =
824064
| | | | | | | | | | | | | | 1 start < =
823487: cytoplasm_(12.15)

 bound_vesicle_ (0.0) | | | | | | | | | | | | | | | | | | | | | NeighGO_term = protein_modification_process_: cytoplasmic_membranebound vesicle (0.0)

| $\begin{aligned} & \text { \| \| } \\ & \text { NeighGO } \end{aligned}$ |
| :---: |
|  |  | (12.26/5.57)

| | | | | | | | | | | | | | |
NeighGO_term = oxidoreductase_activity_: cytoplasmic_membranebound_vesicle_(0.0)
| | | | $\mid$ | | | | | | | | |
NeighGO_term = generation_of_precursor_metabolites_and_energy_:
cytoplasmic_membrane-bound_vesicle_(0.0)
NeighGo_term = cytoplasm_: membrane_(1.35/0.67)
NeighGO_term = transcription_regulator_activity_:
cytoplasmic_membrane-bound_vesicle_ (0.0)
NeighGo_term = signal_transduction_: cytoplasmic_membrane-
bound_vesicle_(0.0)
| | | | | | | | | | | | | | |
NeighGo_term = nucleus_: cytoplasmic_membrane-bound_vesicle_(0.0)
| | | | | | | | | | | | | | | |
NeighGO_term = enzyme_regulator_activity_: cytoplasmic_membrane-
bound vesicle (0.0)
NeighGO_term = ribosome_biogenesis_and_assembly_:
cytoplasmic_membrane-bound_vesicle_ (0.0)
NeighGO_term = ligase_activity_: membrane_(5.12/2.85)
$|\quad| \quad{ }^{-}|\quad| \quad|\quad|^{-}|\quad|-\left.\quad|\quad|\right|^{-}|\quad| \quad|\quad|$
NeighGO_term = translation_: endoplasmic_reticulum_ (7.25/4.3)
| | | | | | | ${ }^{-}$| | | $\mid$| |
NeighGO_term = protein_binding_: cytoplasmic_membrane-bound_vesicle_
(0.0)
| | | | | | | | | | | | | | |
NeighGO_term = transport_: cytoplasmic_membrane-bound_vesicle_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad|$
NeighGO_term = endoplasmic_reticulum_: cytoplasmic_membrane-
bound_vesicle_(0.0)
| $\left.\right|^{-}$| | $\left.\right|^{-}$| | | | | | | | | | |
NeighGo_term = membrane_: cytoplasmic_membrane-bound_vesicle_(0.0)
| ${ }_{\text {NeighGO_term }} \mid$ | $=$ Golgi_apparatus_: cytoplasmic_membrane-bound_vesicle_
(0.0)

NeighGo_term = transporter_activity_: cytoplasmic_membrane-
bound_vesicle_(0.0)
NeighGo_term = hydrolase_activity_: cytoplasmic_membrane-
bound_vesicle_ (0.0)
| | | | | | | | | | | | | | | | NeighGO_term = organelle_organization_and_biogenesis_: membrane_ (4.26/2.48)
| | | | | | | | | | | | | | | NeighGo_term = motor_activity_: cytoplasmic_membrane-bound_vesicle_ (0.0)
| | | | | | | | | | | | | | | | NeighGO_term = cell_cycle_: cytoplasmic_membrane-bound_vesicle_ (0.0)
 cytoplasmic_membrane-bound_vesicle_ (0.0)
| | | | | | | | | $\mid$ | | | | | NeighGO_term = chromosome_: cytoplasmic_membrane-bound_vesicle_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad|$ NeighGO_term = RNA_metabolic_process_: cytoplasmic_membranebound vesicle (0.0 )
| $\left.\right|^{-}$| | ${ }^{-}$| | | | | | | | NeighGO_term = vacuole_: cytoplasmic_membrane-bound_vesicle_ (0.0)

```
NeighGO_term = vesicle-mediated_transport_: cytoplasmic_membrane-
```

bound_vesicle_(0.0)
| $\left.\right|^{-}$| | ${ }^{-}$| | | | 1 | | | | |
NeighGO_term = cytoplasmic_membrane-bound_vesicle_:
cytoplasmic_membrane-bound_vesicle_ (0.0)
NeighGo_term = structural_molecule_activity_: cytoplasmic_membrane-
bound_vēsicle_(0.0)
| | | | $\mid$ | | | | | | | | |
NeighGO_term = peroxisome_: cytoplasmic_membrane-bound_vesicle_
(0.0)
| | | | | | | | | | | | | | |
NeighGO_term = lipid_metabolic_process_: endomembrane_system_
(3.22/1.16)
| | | | | | | | | | | | | | |
NeighGO_term = DNA_binding_: cytoplasmic_membrane-bound_vesicle_
(0.0)
| | | | | | | | | | | | | | |
NeighGO_term = helicase_activity_: cytoplasmic_membrane-
bound vēsicle_(0.0)

NeighGo_term = DNA_metabolic_process_: cytoplasmic_membrane-
bound vēsicle (0. $\overline{0}$ )
| $\left.\right|^{-}$| $\left.\right|^{-}$| | | | | | | | | |
NeighGo_term = meiosis_: cytoplasmic_membrane-bound_vesicle_ (0.0)
$|\quad|-|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ | | |
NeighGO_term = phosphoprotein_phosphatase_activity_:
cytoplasmic_membrane-bound_vesicle_ (0.0)
| | | $\mid$ | | | ${ }^{-}$| ${ }^{-}$| | | | | | |
NeighGO_term = pseudohyphal_growth_: cytoplasmic_membrane-
bound_vesicle_ (0.0)
| | | | | | | | | | | | | | | | NeighGo_term = transcription_: cytoplasmic_membrane-bound_vesicle_ (0.0)
| | | | | | | | | | | | | | | NeighGo_term = protein_kinase_activity_: cytoplasmic_membranebound vesicle (0.0)

```
NeighGO_term = RNA_binding_: cytoplasmic_membrane-bound_vesicle_
```

(0.0)
| | | | | | | | | | | | | | |
NeighGo_term = cell_wall_organization_and_biogenesis_:
cytoplasmic_membrane-bouñ_vesicle_( $\overline{0} .0$ )
| | |
NeighGO_term = cytokinesis_: cytoplasmic_membrane-bound_vesicle_
(0.0)
| | | | | | | | | | | | | | |
NeighGO_term = endomembrane_system_: endomembrane_system_
(6.72/3.13)
| | | | | | | | | | | | | | |
NeighGO_term = nuclear_organization_and_biogenesis_:
cytoplas̄mic_membrane-bōund_vesicle_- (0. $\overline{0}$ )
NeighGO_term = microtubule_organizing_center_: cytoplasmic_membrane-
bound_vesicle_(0.0)
| | | | | | | | | | | | | | |
NeighGO_term = sporulation_: cytoplasmic_membrane-bound_vesicle_
(0.0)
| | | | | | | | | | | | | | |
NeighGo_term = cellular_respiration_: cytoplasmic_membrane-
bound_vesicle_(0.0)
| $\left.\right|^{-}$| | ${ }^{-}$| | | | | 1 | | | |
NeighGO_term = ribosome_: cytoplasmic_membrane-bound_vesicle_ (0.0)
| | $|\quad| \quad|\quad| \quad-|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ | |
NeighGO_term = lyase_activity_: endomembrane_system_(13.75/8.23)

NeighGO_term = amino_acid_and_derivative_metabolic_process_:
cytoplasmic_membrane-bound_vesicle_ (0.0)
NeighGo_term = nucleotidyltransferase_activity_:
cytoplasmic_membrane-bound_vesicle_( $\overline{0} .0$ )
| | | | | | | | | | | | | | |
NeighGO_term = nucleolus_: cytoplasmic_membrane-bound_vesicle_(0.0)

NeighGO_term = cellular_bud_: cytoplasmic_membrane-bound_vesicle_
(0.0)
| | | | | | | | | | | | | | |
NeighGO_term = site_of_polarized_growth_: cytoplasmic_membrane-
bound_vēsicle_ (0.0)
| | | | | | | | | | | | | | |
NeighGO_term = extracellular_region_: cytoplasmic_membrane-
bound_vesicle_ (0.0)
| | | | | | | | | | | | | | |
NeighGO_term = conjugation_: cytoplasmic_membrane-bound_vesicle_
(0.0)
| | | | | | | | | | | | | | | NeighGO_term $=$ response_to_stress_: cytoplasmic_membranebound_vesicle_ (0.0)
| | | | | | | | | | | | | | | NeighGO_term = plasma_membrane_: cytoplasmic_membrane-bound_vesicle_ (0.0)
| | | | | | | | | | | | | | |
NeighGO_term = isomerase_activity_: cytoplasmic_membranebound_vesicle_ (0.0)
| ${ }^{-}$| $\left.\right|^{-}$| | | | | | | | | | NeighGO_term = cell_budding_: cytoplasmic_membrane-bound_vesicle_ (0.0)
| | | | | | | | | | | | | | |
NeighGO_term = translation_regulator_activity_:
cytoplasmic_membrane-bound_vesicle_ (0.0)
NeighGo_term = membrane_fraction_: cytoplasmic_membranebound vēsicle (0.0)
| $\left.\right|^{-}$| $\left.\right|^{-}$| | | | | | | | | | |
NeighGO_term = cell_wall_: cytoplasmic_membrane-bound_vesicle_(0.0)
NeighGO_term = peptidase_activity_: cytoplasmic_membranebound_vesicle_(0.0)
| | | | | | | | | | | | | | | | NeighGo_term = protein_catabolic_process_: cytoplasmic_membranebound_vesicle_(0.0)

NeighGO term = mitochondrial envelope : cytoplasmic membranebound vesicle (0.0)
| | | | | | | | | | | | | | | | NeighGO_term = vitamin_metabolic_process_: cytoplasmic_membranebound_vesicle_(0.0)
NeighGo_term = signal_transducer_activity_: cytoplasmic_membranebound_vesicle_(0.0)
| | | | | | | | | | | | | | | NeighGO_term = membrane_organization_and_biogenesis_: cytoplasmic_membrane-bound_vesicle_ (0.0)
| | $\quad$ | | | | ${ }^{-}$| ${ }^{-}$| | | | | | |
NeighGO_term = carbohydrate_metabolic_process_:
cytoplasmic_membrane-bound_vesicle_ ( 0.0 )
| | | | | | | | | | | | | | | |
NeighGo_term = anatomical_structure_morphogenesis_:
cytoplasmic_membrane-bound_vesicle_-(0.0)
| | | ${ }^{-}$| | $\left.\right|^{-}$| ${ }^{-}$| | | | | NeighGO_term = cell_cortex_: cytoplasmic_membrane-bound_vesicle_ (0.0)
| | | | | | | | | | | | | | | NeighGO_term = cytoskeleton_: cytoplasmic_membrane-bound_vesicle_ (0.0)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>843596$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | strand $=W$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| start $<=$ |  |  |  |  |  |  |  |  |  |  |  |  |  |

845654: cytoplasm_(29.4)







## For Molecular Function






| | | | | | | | | | | start > 207038:
oxidoreductase_activity_ (19.26)
hydrolase_activity_ (17.89)

molecular_function_(9.42)
| | | | | | | | $\left.\right|_{\text {| }}$
(37.63/17.02)
| | | | | | strand = C
| | | | | | | start <= 217529:
transferase_activity ${ }_{\mid}$(12.98)
(17.77)
| | | $\quad$ | $\quad \mid \quad$ start $>226404$
| | | start $<=277623$
| | | | | | start $<=253864$
| | | | | | start $<=235963$

structural molecule activity (13.58)
| | | ${ }^{-}\left|{ }^{-}\right| \quad|\quad| \quad \mid$ start > 228334:
molecular_function_(34.06)
| | | | | | | $\left.\right|^{-}$start > 230248
| | | | | | | | start $<=$ 234512:
hydrolase_activity_ (13.44)
| | | | | | | | start > 234512:
translation_regulator_activity_ (15.93)
| | | | | | | stāt > 235963
| | | | | | | start $<=238124$ :
transferase_activity_ (11.93)
| | | - | | | | start > 238124
| | | | | | | | start $<=249874$
| | | | | | | | | | strand = W

molecular_function_(11.07)
| | | | | ${ }^{-}$| | | | | | start > 238358:
transporter_activity_ (21.62)
| | | | | | | | | | | start > 245721:
molecular_function_(15.53)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ strand $=C$ :
molecular_function_(33.04)
| | | | | | | | start > 249874:
transferase_activity_ (9.13)
| | | | | | start > 253864
| | | | | | | start $<=269194$




| | | | | | | | | | | NeighGO_term = plasma_membrane_: DNA_binding_ (0.0)
| $\left.\right|^{-}|\quad|{ }^{-}| |^{-}| |^{-}|\quad| \quad \mid \quad$ NeighGo_term $=$ isomerase_activity_: DNA_binding_ (0.0)
 cell_budding_: DNA_binding_ (0.0)
| $\mid$ | $\left.\left|{ }^{-}\right|\right|^{-} \mid$| | NeighGO_term = translation_regulator_activity_: DNA_binding_ (0.0)
| | | | | | | | | | | | NeighGo_term = membrane_fraction_: DNA_binding_ (0.0)
 cell_wall_: DNA_binding_ (0.0)
| ${ }^{-}$| - | $\mid$| | | peptidase_activity_: DNA_binding_(0.0)
| | | | | | | | | | | NeighGo_term = protein_catabolic_process_: DNA_binding_ (0.0)
| $\left.{ }^{-}|\quad| \quad\right|^{-}| |^{-}\left|{ }^{-}\right|{ }^{-} \mid \quad$ NeighGo_term $=$ mitochondrial_envelope_: DNA_binding_(0.0)
| | | | | | $\quad$ | $\quad$ | $\mid$ | NeighGo_term $=$ vitamin_metabolic_process_: DNA_binding_ (0.0)
| | | | | | | | | | $\mid$ NeighGO_term = signal_transducer_activity_: transcription_regulator_activity_ (9.36/2.08)
| | | | | | | | | $\mid$ NeighGo_term = membrane_organization_and_biogenesis_: DNA_binding_ (0. $\overline{0}$ )
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ | $\mid$ ighGo_term $=$ carbohydrate_metabolic_process_: DNA_binding_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGo term $=$ anatomical_structure_morphogenesis_: DNA_binding_ (0.0)
| | | | | $\mid$ | | | | $\mid$ | NeighGo_term = cell_cortex_: DNA_binding_ (0.0)
| $|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGo_term $=$ cytoskeleton_: DNA_binding_(0.0)
| | | $\left.\quad\right|^{-}|\quad|{ }^{-}|\quad|$ start > 325334
| | | | | | | | | start <= 343322

molecular_function_(14.43)
| | $\left.\right|^{-}$| $\left.\right|^{-}$| | | | | | | start > 330610 $||||||||||\mid$ start $|=334889$ :
 RNA_binding_(35.99/17.08)
| | | | | | | | |
molecular_function_(15.77)
| | $\left.\right|^{-}| |^{-}|\quad| \quad|\quad| \quad \mid \quad$ start > 343322: RNA_binding_
(21.38)

|  | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | strand $=\mathrm{C}$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=337909$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=330610:$ |  |

protein_binding_ (13.41)
| | | | | | | | | start > 330610:
hydrolase_activity_ (14.66)
| | | ${ }^{-}$| | | | | start > 337909


```
| | | | | | start <= 381274
| | | | | | | | start <= 377612:
nucleotidyltransferase_activity_ (11.06)
| | | | | | | | start > 377612:
transferase activity (17.13)
| | | | | | | start > 381274
| | | | | | | | start <= 385199: helicase_activity
(15.85)
| | | | | | | | start > 385199:
structural_molecule_activity_ (11.49)
| | start > 392228
| | | start <= 601665
| | | | start <= 414106
| | | | | start <= 405779
```



```
| | | | | | | start <= 395835: molecular function
(14.09)
| | | | | | | | start > 395835:
oxidoreductase_activity_ (24.36)
| | | | | | | | | | | start > 398631 
molecular_function_(15.64)
| | | | | | | | | start > 402592: protein_binding_
(10.78)
| | | | | | | start > 403690: molecular_function_
(24.25)
| | | | | | strand = C: molecular_function_ (73.65)
| | | | | start > 405779
| | | | | | start <= 411289
| | | | | | start <= 409607:
translation_regulator_activity_(25.86)
| | | - | | | stārt > 409607: peptidase_activity_
(17.3)
| | | | | | start > 411289
| | | | | | | start <= 411555: transferase_activity_
(14.45)
| | | | | | | start > 411555: molecular_function_
(43.08)
| | | | start > 414106
| | | | | start <= 514558
| | | | | start <= 502942
| | | | | start <= 472303
```



```
| | | | | | | | | | start <= 423967
```



```
transcription_regulator_activity_ (6.47)
| | | | | | | | | | | | start > 418891:
molecular_function_(14.53)
| | | | | | | | | | | start > 423098:
oxidoreductase_activity_ (18.3)
```














```
| | | | | | | | | | | | | | NeighGO_term
= cytoplasm_: protein_kinase_activity_ (1.21/0.49)
| | | | | | | | | | | | | | NeighGO_term
= transcription_regulator_activity_: protein_binding_ (1.86/0.5\overline{7})
| | | | | | | | | | | | | | NeighGO_term
= signal_transduction_: protein_binding_ (0.0)
| | \ | | | - | | - | | - | | | | NeighGO_term
= nucleus_: protein_kinase_activity_ (1.55/0.37)
| | | | | | | | | | | | | | NeighGO_term
= enzyme_regulator_activity_: protein_binding_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= ribosome_biogenesis_and_assembly_: protein_binding_ (2.21/0.4\overline{5})
| | | | | | | | | | | | | | NeighGO_term
= ligase_activity_: protein_binding_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= translation_: protein_kinase_activity_ (0.47/0.13)
| | | | | | | | | | | | | | NeighGO_term
= protein_binding_: protein_binding_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= transport_: protein_binding_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= endoplasmic_reticulum_: protein_binding_ (0.0)
| | | | | | - | | | | | | | | | NeighGO_term
= membrane_: protein_binding_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= Golgi_apparatus_: protein_binding_(0.0)
| | | | | | | | | | | | | | NeighGO_term
= transporter_activity_: protein_binding_ (0.0)
| | | | | | | | \ | \ | | | NeighGO_term
= hydrolase_activity_: protein_binding_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= organelle_organization_and_biogenesis_: protein_binding_(0.0)
| | | - | | | \ \ | | - | | | | | NēighGO_term
= motor_activity_: protein_binding_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= cell_cycle_: protein_binding_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= cytoskeleton_organization_and_biogenesis_: protein_binding_(\overline{0}.0)
| | | | | | | | | | | | | | | | NeighGO_term
= chromosome_: protein_binding_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= RNA_metabolic_process_: protein_binding_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= vacuole_: protein_binding_ (0.0)
| | | | | - | | - | | | | | | | NeighGO_term
= vesicle-mediated_transport_: protein_binding_ (0.0)
= cytoplasmic_membrane-bound_vesicle_: protein_binding_(0.0)
| | | | | | | | | | | | | | NeighGO_term
= structural_molecule_activity_: protein_binding_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= peroxisome_: protein_binding_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= lipid_metabolic_process_: protein_binding_ (0.0)
```

```
| | | | | | | | | | | | | | NeighGO_term
= DNA_binding_: protein_binding_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= helicase_activity_: protein_binding_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= DNA_metabolic_process_: protein_binding_ (0.0)
| | | | - | | - | | | | | |- | | | NeighGO_term
= meiosis_: protein_binding_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= phosphoprotein_phosphatase_activity_: protein_binding_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= pseudohyphal_growth_: protein_binding_ (0.0)
| | | | | | | | - | | | | | | | NeighGO_term
= transcription_: protein_kinase_activity_ (3.33/1.27)
| | | | | | | | | | | | | | NeighGO_term
= protein_kinase_activity_: protein_binding_ (0.0)
| | | | | | | | | - | | - | | | NeighGO_term
= RNA_binding_: protein_binding_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= cell_wall_organization_and_biogenesis_: protein_binding_
(2.55/0.49)
| | | | | | | | | | | | | | NeighGO_term
= cytokinesis_: protein_binding_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= endomembrane_system_: protein_binding_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= nuclear_organization_and_biogenesis_: protein_binding_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= microtubule_organizing_center_: protein_binding_ (0.0)
| | | | | | \ | - | | | | | | NeighGO_term
= sporulation_: protein_binding_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= cellular_respiration_: protein_binding_ (0.0)
| | | | | | - | | \ | \ | | | NeighGO_term
= ribosome_: protein_binding_ (0.0)
| | | | | ` | \ | | | | | | NeighGO_term
= lyase_activity_: protein_binding_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= amino_acid_and_derivative_metabolic_process_: protein_binding_
(0.0)
| | | | | | | | | | | | | | NeighGO_term
= nucleotidyltransferase_activity_: protein_binding_(0.0)
| | | | | | | | | | | | | | NeighGO_term
= nucleolus_: protein_binding_ (0.0)
| | | | | | | | | | | | | | NeighGO term
= cellular_bud_: protein_binding_ (0.0)
| | | | | | \ | \ | | | | | NeighGO_term
= site_of_polarized_growth_: protein_binding_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= extracellular_region_: protein_binding_ (0.0)
| | | | | | | | \ | \ | | | NeighGO_term
= conjugation_: protein_binding_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= response_to_stress_: protein_binding_ (0.0)
```






## For Biological Process

```
        start <= 343322
            start <= 107508
            | start <= 38780
```




```
| | | | | | | | | | NeighGO term =
ribosome_biogenesis_and_assembly_: protein_modification_process_
(0.0)
| | | | | | | | | | NeighGO_term =
ligase_activity_: protein_modification_process_ (0.0)
| | | | | | | | | | NeighGO_term = translation_:
transport_ (2.25/0.57)
| | | | | | | | | | NeighGO_term =
protein_binding_: protein_modification_process_ (0.0)
| | | | | | | | | | NeighGO_term = transport_:
protein_modification_process_ (0.0)
| | | | | ` | ` | | NeighGO term =
endoplasmic_reticulum_: protein_modification_prōcess_ (0.0)
| | | | | | | | | | NeighGO_term = membrane_:
protein_modification_process_ (0.0)
| | | | | | | | | | NeighGO_term =
Golgi_apparatus_: protein_modification_process__(0.0)
| | | | | | |
transporter_activity_: protein_modification_process_ (0.0)
| | | | | | | | | | NeighGO_term =
hydrolase_activity_: protein_modification_process_ (2.56/0.95)
| | | | | | | | | | NeighGO_term =
organelle_organization_and_biogenesis_:
protein_modification_process_ (0.0)
| | | | | | | | | | NeighGO_term =
motor_activity_: protein_modification_process__(0.0)
| | | | | | | | | | NeighG\overline{G_term = cell_cycle_:}
protein_modification_process_ (0.0)
| | | | | \ | \ | | NeighGO_term =
cytoskeleton_organization_and_biogenesis_:
protein_modification_process_ (0.0)
| | | | | | | | | | NeighGO_term = chromosome_:
protein_modification_process_ (0.0)
| | | | | ` | ` | | NeighGO_term =
RNA_metabolic_process_: protein_modification_process_ (0.0)
| | | | | | | | | | NeighGO_term = vacuole_:
protein_modification_process_ (0.0)
| | | | | | | | | | NeighGO_term = vesicle-
mediated_transport_: protein_modification_process_ (0.0)
| | \ | | | | \ | | NēighGO_term =
cytoplasmic_membrane-bound_vesicle_: protein_modification_process_
(0.0)
| | | | | | | | | | NeighGO_term =
structural_molecule_activity_: protein_modification_process_(0.0)
| | | | | - | | \ | | - NeighGO_term = peroxisome_:
protein_modification_process_ (0.0)
| | | | | \ | \ | | NeighGO_term =
lipid_metabolic_process_: protein_modification_process_ (0.0)
| | | | | | | | | | NeighGO_term = DNA_binding_:
protein_modification_process_ (0.0)
| | | | | ` | ` | | NeighGO_term =
helicase_activity_: protein_modification_process__ (0.0)
| | | | | | | | | | NeighGO_term =
DNA_metabolic_process_: protein_modification_process_ (0.0)
```

｜｜｜｜｜｜｜｜｜｜NeighGO＿term＝meiosis＿： protein＿modification＿process＿（0．0）
।｜－｜｜¡｜$\rceil$｜｜NeighGO＿term＝ phosphoprotein＿phosphatase＿activity＿：protein＿modification＿process＿ （0．0）
｜｜｜｜｜｜｜｜｜｜NeighGO term＝
pseudohyphal＿growth＿：protein＿modification＿procēss＿（0．0）
｜｜｜｜｜｜｜｜｜｜NeighGO＿term＝
transcription＿：protein＿modification＿process＿（ $\overline{0} .0$ ）
｜｜｜｜｜｜｜｜｜｜NeighGO＿term＝
protein＿kinase＿activity＿：protein＿modification＿process＿（0．0）
｜｜－｜－｜｜｜｜${ }^{-}$｜NeighGO＿term＝${ }^{-}$RNA＿binding＿： protein＿modification＿process＿（0．0）
｜｜｜｜｜〕｜ $\mid$｜｜NeighGO＿term＝
cell＿wall＿organization＿and＿biogenesis＿：
protein＿modification＿process＿（0．0）
｜｜－｜｜ $\mid$｜$\rceil$｜｜NeighGO＿term＝cytokinesis＿： protein＿modification＿process（0．0）
｜｜－｜｜¡｜$\rceil$｜｜NeighGO＿term＝
endomembrane＿system＿：protein＿modification＿process＿（0．0）
｜｜｜｜｜｜｜｜｜｜NeighGO＿term＝ nuclear＿organization＿and＿biogenesis＿：protein＿modification＿process＿ （0．0）
｜｜｜｜｜｜｜｜｜｜NeighGO＿term＝
microtubule＿organizing＿center＿：protein＿modification＿process＿（0．0）
$|\quad| \quad|\quad| \quad|\quad| \quad||\mid$ NeighGO＿term $=$ sporulation＿：
protein＿modification＿process＿（0．0）
｜｜－｜｜ $\mid$｜ $\mid$｜｜NeighGO term $=$ cellular＿respiration＿：protein＿modification＿process＿（0．0）
｜｜｜｜｜｜｜｜｜｜NeighGO＿term＝ribosome＿：
protein＿modification＿process＿（0．0）
｜｜｜｜｜〕｜｜｜｜NeighGO＿term＝ lyase＿activity＿：protein＿modification＿process＿（0．0）
।｜ $\mid$｜｜｜$\left.\overline{-}|\quad| \quad\right|^{-}$Neigh $\bar{G} O \_$term $=$ amino＿acid＿and＿derivative＿metabolic＿process＿： protein＿modification＿process＿（0．0）
｜｜｜｜｜〕｜｜｜｜NeighGO＿term＝
nucleotidyltransferase＿activity＿：protein＿modification＿process＿
（0．0）
｜｜｜｜｜｜｜｜｜｜NeighGo＿term＝nucleolus＿：
protein＿modification＿process＿（0．0）
｜｜｜｜｜｜｜｜｜｜NeighGO＿term＝
cellular＿bud＿：protein＿modification＿process＿（0．0）
।｜$\overline{\mid} \quad|\quad|-\quad|\quad| \quad \mid \quad$｜NeiḡhGO term $=$
site＿of＿polarized＿growth＿：protein＿modification＿process＿（0．0）
｜$\quad$｜｜｜｜ $\mid$｜｜ $\mid$ NeighGO＿term $=$
extracellular＿region＿：protein＿modification＿process＿（0．0）
｜｜｜$\left.\right|^{-}$｜$\rangle$｜$\left.\right|^{-}$｜｜NeighGO＿term＝conjugation＿：
protein＿modification＿process＿（0．0）
｜｜－｜｜¡｜$\rceil$｜｜NeighGO term $=$ response＿to＿stress＿：protein＿modification＿process＿（0．0） ｜｜ $\mid$｜｜｜｜｜｜NeighGO＿term＝ plasma＿membrane＿：protein＿modification＿process＿（0．0）

```
| | | | | | | | | | NeighGO term =
isomerase_activity_: protein modification_processs_(0.0)
| | | | | | | | | | NeighGO_term =
cell_budding_: protein_modification_process_ (0.0)
| | | | | | | | | NeighGO_term =
translation_regulator activity_: protein_modific
| | | - | | | | | | N
membrane_fraction_: protein_modification_process_ (0.0)
| | | | | | | | | | NeighGO_term = cell_wall_:
protein_modification_process_ (0.0)
| | | | | | | | | | NeighGO_term =
peptidase_activity_: protein_modification_process_ (0.0)
| | | | | | | \ | | NēighGO_tērm =
protein_catabolic_process_: protein_modification_process_ (0.0)
| | | | | | | | | | NeighGO_term =
mitochondrial_envelope_: protein_modification_process_ (0.0)
| | | | | | | | | | NeighGO term =
vitamin metabolic process : protein modification process (0.0)
| | | | | | | | | | NeighGO_term =
signal_transducer_activity_: protein_modification_process_ (0.0)
| | | | | | | | | | NeighGO_term =
membrane_organization_and_biogenesis_: protein_modification_process_
(0.0)
| | | | | | | | | | NeighGO_term =
carbohydrate_metabolic_process_: protein_modification_process_ (0.0)
| | | | | | | | | | NeighGO_term =
anatomical_structure_morphogenesis_: protein_modification_process_
(0.0)
| | | | | | | | | NeighGO_term = cell_cortex_:
protein_modification_process_ (0.0)
| | | | | | | | | | NeighGO_term =
cytoskeleton_: protein_modification_process_ (0.0)
| | | | | | | | strand = C
| | | | | | | | start <= 73339:
protein_modification_process_ (19.88)
| | | | | | | | | start > 73339: translation_
(18.15)
| | | | | | | start > 76893
| | | | | | start <= 80364
| | | | | | | | | start <= 78856:
biological_process_ (15.33)
| | | | | | | | | start > 78856: transport_
(35.82/15.34)
| | | | | | | | start > 80364
| | | | | | | | | start <= 81185:
RNA_metabolic_process_ (19.13)
| | | | | | | | | start > 81185:
biological_process_ (38.11)
| | | | start > 82373
| | | | | start <= 91251
| | | | | | start <= 84256:
organelle_organization_and_biogenesis_ (19.62)
| | | | | | start > 84256
```



```
(13.68)
| | | | | | | | | strand = C:
DNA_metabolic_process_(7.39)
|
organelle_organization_and_biogenesis_ (18.26)
| | | | | | - stārt > 12470\overline{3}
| | | | | | start <= 154312
| | | | | | | start <= 142252
| | | | | | | | | start <= 129888:
biological process (14.36)
| | | | | | | | | | start > 129888
| | | | | | | | | | | start <= 131531:
translation_(12.46)
| | | | | | | | | | | start > 131531:
biological_process_ (11.73)
| | | | | | | | | start > 139967
| | | | | | | | | | start <= 141732
| | | | | | | | | | | NeighGO_term =
cellular_component_: meiosis_(13.15/7.24)
| | | | | | ` | | | NeighGO_term =
molecular function
| | | | | | | | | | | | neigh_strand = W:
meiosis_(2.32/1.19)
| | | | | | | | | | | | neigh_strand = C:
DNA_metabolic_process_(5.98/3.18)
| | | | | |
biological_process_
| | | | | | | | | | | distance <= 1112:
DNA_metabolic_process_(3.71/1.75)
| | | | | | | | | | | | distance > 1112:
transcription (6.9/1.94)
| | | | | | | | | | | NeighGO_term =
transferase_activity_: DNA_metabolic_process_(3.8/1.22)
| | | | | | | | | | | NeighGO_term =
protein_modification_process_: transcription_ (0.0)
| | | | | \ | \ | | | N
mitochondrion : meiosis (3.39/1.45)
| | | | | | | | | | | NeighGO_term =
oxidoreductase_activity_: transcription_ (0.0)
| | | | | | | | | | | NeighGO_term =
generation_of_precursor_metabolites_and_energy_: transcription_
(0.0)
| | | | | | | | | | NeighGO_term =
cytoplasm_: transcription_(0.0)
transcription_regulator_activity_: transcription_(0.0)
```











```
| | | | | | | | start <= 392228:
biological_process_ (22.59)
| | | | | | | | | start > 392228:
DNA_metabolic_process_ (13.91)
| | | | | | | | strand = C: biological_process_
(46.48)
| |:| | | | | | | | | start > 397624 
organelle_organization_and_biogenesis_ (26.25)
| | | | | | | | | start > 398631
| | | | | | | | | start <= 402592:
conjugation_(13.1)
| | | | | | | | | | start > 402592:
organelle_organization_and_biogenesis_ (12.41)
```



```
(17.11)
| | | | | | | | | start > 401290: translation_
(19.78)
\begin{tabular}{lllll}
1 & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) \\
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) \\
1 & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) \\
1 & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) \\
1 & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) \\
1 & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\)
\end{tabular}
| | | - | | | | | | | start > 405473:
RNA metabolic process (20.8)
| | | | | | | | | | start > 405779:
protein_modification_process_ (20.73)
```



```
| | | | | | | | | | start <= 409607:
translation_(9.75)
| | | | | | | | | | | start > 409607:
protein_catabolic_process_ (29.15)
| | | | | | | | | start > 411555: translation
(24.22)
| | | | | | | | | | | | | start > 412370 
(24.37)
| | | | | | | | | start > 414106:
RNA_metabolic_process_(65.66/34.49)
| - | | | | | | start > 417487
| | | | | | | start <= 428606
| | | | | | | | start <= 426810
```



```
translation_(11.22)
| | | | | | | | | | | start > 423098:
DNA_metabolic_process_ (11.79)
```

```
| | | | | | | | | | | start <= 423098:
biological_process_ (18.19)
| | | | | | | | | | | start > 423098:
protein_modification_process_ (16.45)
| | | | | \ | \ | start > 426810
| | | | | | | | | start <= 427953:
vitamin metabolic process (27.2/12.68)
| | | | | | | | | | start > 427953:
biological_process_ (31.43)
| | | | | | | | start > 428606
| | | | | | | | | | strand = W:
DNA metabolic process (28.24/16.79)
```



```
transcription (16.65)
| | | | | | | | | | | start > 433067:
conjugation_(14.1)
\begin{tabular}{llllllllll}
\(\mid\) & \(\mid\) & \(\mid\) & - & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & start \(>436838\) \\
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & strand \(=W\) \\
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(|\quad| \quad\) start \(<=439096:\)
\end{tabular}
translation_(4.82)
| | | | | | | | | | | start > 439096:
ribosome_biogenesis_and_assembly_ (15.57)
| | | | | | | | | | strand = C: translation_
(31.96)
| | | start > 440812
| | | | start <= 527333
| | | | | start <= 459859
| | | | | | start <= 452109
| | | | | | start <= 448540
| | | | | | | start <= 446148
start <= 442914: cytokinesis
(9.28)
| | | | | | | | | start > 442914:
cell_wall_organization_and_biogenesis_ (14.73)
| | | | | | | | start > 446148:
amino_acid_and_derivative_metabolic_process_ (23.68)
| | | | | | | | start > - 448540-
| | | | | | | | start <= 449962: transcription_
(35.43/17.28)
| | | | | | | | start > 449962: vesicle-
mediated_transport_ (27.1)
| | | | | | start > 452109
| | | | | | start <= 458161
| | | | | | | | start <= 454790:
protein_modification_process_ (40.36)
| | | | | | | | start > 454790
| | | | | | | | | start <= 457875:
organelle_organization_and_biogenesis_(52.49/26.67)
| | | | | | | | | start > 457875:
response_to_stress_ (15.63)
```



| | | | | | | | | | | NeighGO_term = enzyme_regulator_activity_:
amino_acid_and_derivative_metabolic_process_ (0.0)
| $\left.\left.\left.\left.\right|^{-}\right|^{-}\right|^{-}|\quad|\right|^{-}|\quad| \quad{ }^{-} \mid \quad$ | NeighGo_term =
ribosome_biogenesis_and_assembly_:
amino_acīd_and_derivativ̄e_metabolic_process_ (0.0)
| $\left.\left.\left.\left.\right|^{-}\right|^{-}\right|^{-}|\quad|\right|^{-}|\quad| \quad{ }^{-} \mid \quad$ NeighGo_term $=$ ligase_activity_: amino_acid_and_derivative_metabolic_process_(0.0)
| | | | | | | | | | | NeighGO_term =
translation_: amino_acid_and_derivative_metabolic_process_ (0.0)
 protein_binding_: amino_acid_and_derivative_metabolic_process_(0.0) | | - | - | - $\quad$ | $\overline{\text { | }}$ | ${ }^{-}$| Neigh $\bar{G} O$ term $\overline{=}$ transport_: amino_acid_and_derivative_metabolic_process_ (0.0) $|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGO_term $=$ endoplasmic_reticulum_: amino_acid_and_derivative_metabolic_process_ (0.0)
| | | | | | | | | NeighGO_term = membrane_: lipid_metabolic_process_(6.35/2.38)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ | $|\quad| \quad$ lighGo_term $=$ Golgi_apparatus_: amino_acid_and_derivative_metabolic_process_(0.0)
| | | | | | | | | | | | NeighGO_term = transporter_activity_: amino_acid_and_derivative_metabolic_process_ (0.0)
| | | | | | | | | | hydrolase_activity_: amino_acid_and_derivative_metabolic_process_ (0.0)
| | | | | | | | | | | NeighGo_term = organelle_organization_and_biogenesis_:
amino_acid_and_derivative_metabolic_process_ (0.0)
| | $\left.\left.{ }^{-}\right|^{-}|\quad|\right|^{-}|\quad| \quad|\quad| \quad$ NeighGo_term $=$
motor_activity_: amino_acid_and_derivative_metabolic_process_(0.0)
 cell_cycle_: amino_acid_and_derivative_metabolic_proces $\bar{s}$ ( 0.0 ) | $\mid$ | | | | $\mid$ | | $\mid$ | cytoskeleton_organization_and_biogenesis_: amino_acid_and_derivative_metabolic_process_ (0.0)
| |- | | | | | | | | | chromosome_: amino_acid_and_derivative_metabolic_process _ (0.0) $|\quad| \quad|\quad| \quad\left|-\left|\left.\right|^{-}\right| \quad\right| \quad \mid \quad$ NeighGo_term = RNA_metabolic_process_: amino_acid_and_derivative_metabolic_process_ (3.43/1.73)
| | | | | | | | | | | NeighGO_term = vacuole_: amino_acid_and_derivative_metabolic_process_(0.0) $|\quad| \quad|\quad| \quad \mid \quad$ | $|\quad| \quad$ | vesicle-mediated_transport_:
amino_acid_and_derivative_metabolic_process_ (0.0) cytoplasmic_membrane-bound_vesicle_: amino_acid_̄̄and_derivative_Metabolic__process_ (0.0) | $\left.\left.\left.\right|^{-}\right|^{-}|\quad|\right|^{-}|\quad| \quad \mid \quad$ - $\mid$ NeighGO_term $=$ structural_molecule_activity_:
amino_acid_and_derivative_metabolic_process_ (0.0)
| | | | | | | | | | | | NeighGO term = peroxisome_: amino_acid_and_derivative_metabolic_process̄ (0.0) $|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGo_term $=$ lipid_metabolic_process_:
amino_acid_and_derivative_metabolic_process_ (0.0)
| $\left.\left.\left.\right|^{-}\right|^{-}|\quad|\right|^{-}|\quad| \quad|\quad-|$ NeighGO_term $=$ DNA_binding_: amino_acid_and_derivative_metabolic_process (0.0) | | | | | | $\mid$ | | $\mid$ | NeighGo_tērm = helicase_activity_: amino_acid_and_derivative_metabolic_process_ (0.0)
| | | | | | | | | | | | NeighGO_term = DNA_metabolic_process_: amino_acid_and_derivative_metabōic_process (0.0)
| | | | | | | | | | | | NeighGO_term = meiosis_: amino_acid_and_derivative_metabolic_process_ (0.0)
 phosphoprotein_phosphatase_activity_:
amino_acid_and_derivative_metabolic_process_ (0.0)
| $\left.\left.\left.\left.\right|^{-}\right|^{-}\right|^{-}|\quad|\right|^{-}|\quad|{ }^{-}|\quad| \quad$ | NeighGo_term $=$ pseudohyphal_growth_: amino_acid_and_derivative_metabolic_process_ (0.0)
 transcription_: lipid_metabolic_process_ (3.76/1.93) | | | ${ }^{-}$| ${ }^{-}$| | - | ${ }^{-}$| NeighGo_term = protein_kinase_activity_:
amino_acid_and_derivative_metabolic_process_ (0.0)
$\left.\left.\left.\left|\left.\right|^{-}\right|^{-}\right|^{-}|\quad|\right|^{-}|\quad| \quad\right|_{\mid} ^{-} \mid \quad$ NeighGo_term = RNA_binding_: biological_process_(4.97/1.8)
| $\mid$ | | | $\mid$ | $\mid$ | | NeighGo_term $=$ cell_wall_organization_and_biogenesis_: amino_acid_and_derivative_metabolic_process_ (0.0)
| $\left.\left.\left.\left.\right|^{-}\right|^{-}\right|^{-}|\quad|\right|^{-}|\quad| \quad{ }^{-} \mid \quad{ }^{-}$| NeighGo_term = cytokinesis_: amino_acid_and_derivative_metabolic_processs_(0.0)
 endomembrane_system_: amino_acid_and_derivative_metabolīc_process_ (0.0)

nuclear_organization_and_biogenesis_:
amino_a $\bar{c} i d$ and_derivātive_metabolic_process_ (0.0)
| | $\left.\left.{ }^{-}\right|^{-}|\quad|\right|^{-}|\quad| \quad \mid \quad$ | ${ }^{-}$| NeighGo_term = microtubule_organizing_center_: amino_acid_and_derivative_metabolic_process_ (0.0)
$\left.\left|\left.\right|^{-}\right|^{-}\right|^{-}| |^{-}|\quad|{ }^{-}|\quad| \quad{ }^{-} \mid \quad$ NeighGo_term = sporulation_: amino_acid_and_derivative_metabolic_process_ (0.0) $|\quad| \quad$ - $|\quad| \quad$ | $\quad$ | $\mid \quad$ | cellular_respiration_: amino_acid_and_derivative_metabolic_process_ (0.0)
| | | | | | | | | | |
ribosome_: amino_acid_and_derivative_metabolic_process_-(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad{ }^{-}$NeighḠ - $\quad \mid$ erm $=$ lyase_activity_: amino_acid_and_derivative_metabolic_process_ (0.0) | | | | | | | | | | | NeighGO_term =
amino_acid_and_derivative_metabolic_process_:
amino_acid_and_derivative_metabolic_process_ (0.0)
| | | | | | | | | | | NeighGO_term = nucleotidyltransferase_activity_:
amino_acid_and_derivative_metabolic_process_ (0.0)
$\left.\left|\left.\right|^{-}\right|^{-}|\quad|\right|^{-}|\quad| \quad|\quad| \quad$ NeighGo_term $=$ nucleolus_: amino_acid_and_derivative_metabolic_process_ (0.0) | | | | | | | | | ${ }^{-}$NeighGO_term = cellular_bud_: amino_acid_and_derivative_metabolic_procēss_ (0.0)
 site_of_polarized_growth_: amino_acid_and_derivative_metabolic_process_ (0.0)
$\left.\left|\left.\right|^{-}\right|\right|_{\mid}|\quad| \quad|\quad| \quad \mid \quad$ | NeighGo term = extracellular_region_: amino_acid_and_derivative_metaboliic_process_ (0.0)
| | | | | | | | | | NeighGO_term = conjugation_: amino_acid_and_derivative_metabolic_process_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGO_term $=$ response_to_stress_: amino_acid_and_derivative_metabolic_process_ (4.77/0.81)
| | | | | | | | | $\mid$ NeighGo_term = plasma_membrane_: amino_acid_and_derivative_metabolic_process_ (0.0) | | | | | | | | | | | | NeighGO_term = isomerase_activity_: amino_acid_and_derivative_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO_term = cell_budding_: amino_acid_and_derivative_metabolic_process_ (0.0)
 translation_regulator_activity_:

| $\left.\left.\left.\left.\right|^{-}\right|^{-}\right|^{-}|\quad|\right|^{-}|\quad| \quad\left|\quad{ }^{-}\right| \quad$ NeighGo_term $=$ membrane_fraction_: amino_acid_and_derivative_metabolic_process_ (0.0)
| | | | | | | | | | | | NeighGO_term = cell_wall_: amino_acid_and_derivative_metabolic_process_ (0.0) | $\left.\left.\left.\left.\quad\right|^{-}|\quad|^{-}\right|^{-}\right|^{-}|\quad| \quad\right|^{-}|\quad| \quad$ NeighGO_term = peptidase_activity_: amino_acid_and_derivative_metabolic_process_ (0.0)
| | | | | | | | | | | protein_catabolic_process_: amino_a $\bar{c} i d \_a n d \_d e \bar{r} i v a t i v e$ - metabolic process (0.0) | $\left.\left.\left.\left.\right|^{-}\right|^{-}\right|^{-}|\quad|\right|^{-}|\quad| \quad{ }^{-}\left|{ }^{-}\right| \quad$ NeighGo_term $=$ mitochondrial_envelope_:
amino_acid_and_derivative_metabolic_process_ (4.1/1.7)
$\left.\left.\left|\left.\right|^{-}\right|^{-}\right|^{-}|\quad|\right|^{-}|\quad|{ }^{-}|\quad| \quad{ }^{-} \mid \quad$ NeighGo_term = vitamin_metabolic_process_:
amino_acid_and_derivative_metabolic_process_ (0.0)
| | $\left.\left.{ }^{-}\right|^{-}|\quad|\right|^{-}|\quad| \quad|\quad| \quad{ }^{-} \mid \quad$ NeighGo_term $=$
signal_transducer_activity_:
amino_acid_and_derivative_metabolic_process_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ | $\mid$ | $\mid$ ighGo_term $=$ membrane_organization_and_biogenesis_:
amino_acid_and_derivative_metabolic_process_ (0.0)
| | $\left.\left.{ }^{-}\right|^{-}|\quad|\right|^{-}|\quad| \quad|\quad| \quad$ | NeighGo_term $=$
carbohydrate_metabolic_process_:
amino_acid_and_derivative_metabolic_process_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGo_term $=$ anatomical_structure_morphogenesis_:
amino_acid_and_derivative_metabolic_process_ (0.0)


cytoskeleton_: amino_acid_and_derivative_metabolic_procēss_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ | $\mid \quad$ start $>508368$ :
vitamin_metabolic_process_(4.46)
$\begin{array}{llllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }>511549 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=523791\end{array}$
$\mid$ | | | | | | | | | | start $<=516947$
| | | | | | | | | | | | start <= 513163:
biological_process_(9.36)

867: response_to_stress_ (2.46/1.15)
| | | $\quad$ | $\mid$ | | | | | | | distance >
867: RNA_metabolic_process_ (5.56)
| | | | | | | | | | | | start > 516947:
biological_process_(22.5)
| | | | | | | | | | start > 523791:
amino_acid_and_derivative_metabolic_process_(27.78/11.01)
$\begin{array}{llllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { distance }>2663 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=514558 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\ \text { vitart }<=513163 \\ \text { vitamin_metabolic_process_(3.47) }\end{array}$
| | | | | | | | | | | | start > 511549:
biological_process_ (9.62)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ start > 513163:
RNA_metabolic_process_ (20.88)

| - | $\mid$ | $\mid$ | $\left.\right\|^{-}$ | $\mid$ | $\mid$ | $\mid$ |
| ---: | ---: | ---: | ---: | ---: | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>514558$ |
| $\mid$ | $\mid$ | $\mid$ | start $<=516947:$ |  |  |  |

signal_transduction_(28.0/12.3)

biological_process_ (5.1)
| | | | | | | | strand = C
| | | | | | | start $<=514558$
| | | | | | | | start <= 508368:
response_to_stress_ (17.16)
$|\quad| \quad|\quad| \quad|\quad| \quad$ start > 508368: vesicle-
mediated_transport_ (12.91)
$|\quad| \quad|\quad|-|\quad| \quad|\quad| \quad$ start $>514558$
| | | | | | | | | start <= 523791:
biological_process_ (19.69)
| | | | | | | | | | start > 523791:
organelle_organization_and_biogenesis_ (16.74)
| | | ${ }^{-}$| start $>52 \overline{7} 333$
| | | | | start $<=676625$
| | | | | strand $=\mathrm{W}$
| | | | | | start $<=601665$
$|||||\mid \quad$ start $<=561747$






```
| | | | | | | | | | start > 731142:
biological_process_ (8.37)
| | | | | | | | | start > 733417
| | | | | | | | | | start <= 733940:
transcription (12.13)
| | | | | | | | | start > 733940: translation
(23.76)
| | | | | | | | start > 739127:
amino_acid_and_derivative_metabolic_process_ (19.13)
| | | | | | start > 739949 - 
| | | | | | | | start <= 754731
| | | | | | | | | start <= 746803:
biological_process_ (51.27)
| | | | | | | | | | start > 746803
| | | | | | | | | | | start <= 750405:
cell_cycle_(15.17)
| \ | | | | | | | | | start > 750405:
biological_process_ (11.21)
| | | | | | | | | start > 754731
```



```
organelle organization and biogenesis (15.57)
| | | | | | | | | | | start > 756901:
transcription_(23.65)
| | | | | | | | | | start > 757775:
biological_process_ (39.01)
| | | | | | | | start > 767434: transport
(34.76/8.68)
\begin{tabular}{lllllll}
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & strand \(=C\) \\
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & start \(<=731142\) \\
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) start \(<=729676\) \\
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\)
\end{tabular}
biological_process_ (17.78)
| | | | | | | | | start > 726978:
nuclear_organization_and_biogenesis_(19.36)
| | | | | | | | start > 729676: transport_
(32.99/8.72)
| | | | | | start > 731142
| | | | | | | start <= 750405
| | | | | | | | start <= 733940:
biological_process_ (10.46)
| | | | | | | | | start > 733940:
ribosome_biogenesis_and_assembly_(26.49)
| | | | | - | | | start > 750405
| | | | | | | | | start <= 757775:
biological_process_ (38.68)
| | | | | | | | | start > 757775: transport_
(10.05)
| | | | | start > 772459
| | | | | | start <= 781772
| | | | | | | start <= 779621
```




| | | | | | | | | | start > 885746:
carbohydrate_metabolic_process_(11.2)
$|||\quad| \quad| \quad| \quad \mid \quad$ start $>888887$ :
biological_process_(40.34)
| | | | | | | | start > 897507
| | | | | | | | start <= 902275: RNA_metabolic_process (13.23)

| $\mid$ | - | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>902275$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=905939$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| biological_process__ | $(20.38)$ |  |  |  |  |  |  |  |  |

| | | | | | | | | | start > 903202:
lipid_metabolic_process_ (12.8)
| | | | | | | | | | start > 905939:
biological_process_ (35.27)
| | | | | | start > 911633:
amino_acid_and_derivative_metabolic_process_ (26.97)
| |- | ${ }^{-}$stārt > 911888
| | | | start $<=965660$
| | | | | start $<=947423$
| | | |
| | | | | | | start <= 913232: response_to_stress_
(18.71)

response_to_stress_ (16.53)
| | | | | | | start > 915246: biological_process_
(24.62)
| | | | | | start > 919470
| | | | | | start $<=939926$
| | | | | | | start $<=936886$
| | | | | | | | | strand = W
| | | | | | | | | start <= 922180:
translation_ (26.77)
$\begin{array}{llllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }>\text { 922180 } \\ \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=924699:\end{array}$
transport_ (12.63)
| | | | | | | | | | start > 924699:
ribosome_biogenesis_and_assembly_ (3.94)
| | $\quad|\quad| \quad$ | - | $\quad$ | strand $=\mathrm{C}$
| | | | | | | | | start <= 924699:
protein_modification_process_(16.5)
| | | | | $\mid$ | $|\quad| \quad \mid \quad$ start > 924699: translation (23.47)



```
| | | | | | | | | | NeighGO_term =
transcription_regulator_activity_
organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | NeighGO_term =
signal_transduction_: organelle_organization_and_biogenesis_(0.0)
| | | | | | | | | | NeighGO_term = nucleus_:
organelle organization and biogenesis (0.0)
| | | | | | | | | | NeighGO_term =
enzyme_regulator_activity_: organelle_organization_and_biogenesis_
(0.0)
| | | | | | | | | | NeighGO_term =
ribosome biogenesis_and assembly_
organelle_organization__and_biogenesis_ (0.0)
| | | | | | | | | | NeighGO_term =
ligase_activity_: organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | NeighGO_term = translation_:
organelle_organization_and_biogenesis_ (0.0)
| | | | | | - | | | |
protein_binding_: organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | NeighGO_term = transport_:
organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | NeighGO term =
endoplasmic_reticulum_: organelle_organization ānd biogenesis (0.0)
| | | | | | | | | | NeighGO_term = membrañe_:
organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | NeighGO_term =
Golgi_apparatus_: organelle_organization_and_biogenesis_ (0.0)
| | | | - | | | - | | | N
transporter_activity_: organelle_organization_añd_biogenesis_ (0.0)
| | | | | | | | | | NeighGO_term =
hydrolase_activity_: organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | NeighGO_term =
organelle_organization_and_biogenesis_:
organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | NeighGO_term =
motor_activity_: organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | NeighGO_term = cell_cycle_:
organelle_organization_and_biogenesis_ (0.0)
| | | | | | - | | | | NeighGO_term =
cytoskeleton_organization_and_biogenesis_:
organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | NeighGO_term = chromosome_:
organelle_organization_and_biogenesis_ (0.0)
| | |
RNA_metabolic_process_: organelle_organization_ānd_biogenesis_(0.0)
| | | | | | | | | | NeighGO_term = vacuole_:
organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | NeighGO_term = vesicle-
mediated_transport_: organelle_organization_and_biogenesis_ (0.0)
| | \ | | | | | - | | Nei\overline{ghGO_term =}
cytoplasmic_membrane-bound_vesicle_:
organelle_organization_and_biogenesis_ (0.0)
```

```
| | | | | | | | | | NeighGO_term =
structural molecule_activity :
organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | NeighGO_term = peroxisome_:
organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | NeighGO term =
lipid_metabolic_process_: organelle_organization_and_biogenesis
(0.0)
| | | | | | | | | | NeighGO_term = DNA_binding_:
organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | NeighGO term =
helicase_activity_: organelle_organization_and_\overline{b}iogenesis_ (0.0)
| | | | | | | | | | NeighGO_term =
DNA_metabolic_process_: organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | NeighGO_term = meiosis_:
organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | NeighGO_term =
phosphoprotein_phosphatase_activity_:
organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | NeighGO_term =
pseudohyphal_growth_: organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | NeighGO_term =
transcription : organelle organization and biogēnesis (0.0)
| | | | | | | | | | NeighGO_term =
protein_kinase_activity_: organelle_organization_and_biogenesis_
(0.0)
| | | | | | | | | | NeighGO_term = RNA_binding_:
organelle_organization_and_biogenesis_ (0.0)
| | | | | | - | | | | NeighGO_term =
cell_wall_organization_and_biogenesis_:
organelle_organization_and_biogenesis__ (0.0)
| | | | | | | | | | NeighGO_term = cytokinesis_:
organelle_organization_and_biogenesis_ (0.0)
| | | | | | - | | | |
endomembrane_system_: organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | NeighGO_term =
nuclear_organization_and_biogenesis_:
organel\overline{l}e_organization_a\overline{n}d_biogenesis_ (0.0)
| | | | | | - | - | | | | NeighGO_term =
microtubule_organizing_center_:
organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | NeighGO_term = sporulation_:
organelle_organization_and_biogenesis_ (0.0)
| | | | | | - | - | | | | NeighGO term =
cellular_respiration_: organelle_organization_añd_biogenesis_(0.0)
| | | | | | | | | | NeighGO_term = ribosome_:
organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | NeighGO_term =
lyase_activity_: organelle_organization_and_biogenesis_ (0.0)
| | | | | | | - | | | - Neig}hGO_term =
amino_acid_and_derivative_metabolic_process_:
organelle_organization_and_biogenesis_ (0.0)
```

| | | | | | | | | NeighGO_term = nucleotidyltransferase_activity_:
organelle_organization_and_biogenesis_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGO_term = nucleolus_:
organelle_organization_and_biogenesis_ (0.0)

cellular_bud_: transport (2.82/0.69)
$|\quad| \quad|\quad| \quad \mid \quad$ | $\mid$ NeighGO_term $=$
site_of_polarized_growth_: organelle_organization_and_biogenesis_
(2.31/1.03)
| | | | | | | | | NeighGO_term =
extracellular_region_: organelle_organization_añ_biogenesis_(0.0)
$|\quad| \quad|\quad| \quad \overline{\mid}|\quad| \quad \mid \quad$ Neigh $\bar{G} O \_$tērm $=$conjuḡation_:
transport_(4.27/0.77)
| | | | | | | | | NeighGO_term =
response_to_stress_: organelle_organization_and_biogenesis_
(4.69/1. $\overline{2} 2$ )
| | | | | | | | $\mid$ NeighGO_term =
plasma_membrane_: organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | 1 NeighGO_term =
isomerase_activity_: organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | NeighGO_term =
cell_budding_: organelle_organization_and_biogeñesis_(0.0)
| $\mid$ | $\mid$ | $\mid$ | | ${ }^{-}$NeighGO_term ${ }^{-}=$
translation_regulator_activity_:
organelle_organization_and_biogenesis_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGO term =
membrane fraction_: organelle_organization_and_而iogenesis (0.0)

organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | NeighGO_term =
peptidase_activity_: organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | NeighGOterm =
protein_catabolic_process_: organelle_organization_and_biogenesis_
(0.0)
| | | | | | | | | NeighGO_term =
mitochondrial_envelope_: organelle_organization_and_biogenesis_
(0.0)
| | | | | | | | NeighGO term =
vitamin_metabolic_process_: organelle_organization_and_biogenesis_
(0.0)
| | | | | | | | | | NeighGO_term =
signal_transducer_activity_: organelle_organization_and_biogenesis_
(0.0)
| | | | | | | | NeighGO_term =
membrane_organization_and_biogenesis_:
organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | NeighGO_term =
carbohydrate_metabolic_process_: transport_(1.58/0.32)
$|\quad| \quad|\quad| \quad|\quad| \quad$ | $\mid \quad$ NeighGO_term $=$
anatomical_structure_morphogenesis_:
organelle_organization_and_biogenesis_ (0.0)
$\left.\left.|\quad|\right|^{-}|\quad|\right|^{-}\left|-|\quad| \quad\right.$ NeighGO_term $=c e l l_{-} c o r t e x \_: ~$
organelle_organization_and_biogenesis_ (0.0)



```
| | | | | | | | start <= 1080313:
biological_process_(7.18)
| | | | | | | | | start > 1080313:
DNA_metabolic_process_ (12.0)
| | | | | | | strand = C | | start <= 1060047
| | | | | | | start <= 1049965
```



```
ribosome_biogenesis_and_assembly_ (15.83)
| | | | | | | | | | start > 1045487:
biological_process_ (23.43)
| | | | | | | | | start > 1046738: conjugation_
(20.37)
| | | | | | | | | | start > 1049965 
ribosome_biogenesis_and_assembly_ (21.1)
| | \ | | - | | | \ start > 1052831
| | | | | | | | | | start <= 1058731:
cell_wall_organization_and_biogenesis_ (22.21)
| | | | | | | | | | start > 1058731:
ribosome_biogenesis_and_assembly_ (17.13)
| | | | | - - | start > 1060047
| | | | | | | | start <= 1063160
| | | | | | | | start <= 1061597: transport_
(39.38/18.28)
| | | | | | | | start > 1061597: translation_
(29.18)
| | | | | | | | start > 1063160
| | | | | | | | | start <= 1064947:
vitamin_metabolic_process_ (19.21)
| | | | | | | | | start > 1064947
| | | | | | | | | start <= 1070300:
biological_process_ (21.22)
| | | | | | | | | | start > 1070300
| | | | | | | | | | start <= 1076606:
transport_ (14.08)
| | | | | | | | | | | start > 1076606:
biological_process_ (6.7)
```


## Decision Tree Generated for Chromosome Eight:

## For Cellular Component

| start <= 517532
| | start <= 222480
| | | start $<=148662$
| | | | start $<=74696$
| | | | | start $<=20968$
| | | | | start $<=6400$
| strand = C: cellular_component_ (24.42)
| strand = W: endomembrane_system_ (16.35)
start > 6400
| start <= 15665
| | start $<=10211$


| | | | | | | | | | | start > 78932: cytoplasm_(17.07)
| | | | | | | | | start > 81612:
cellular_component_(14.98)
 (9.68)
| | | | | | | | | start > 94506: site_of_polarized_growth_(15.03)
 (19.19)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | strand $=W$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid \quad$ start $<=$ 99215: mitochondrion_ | (17.38)




```
| | | | | | | | | | NeighGO_term = transport_:
membrane (0.0)
| | \ | | | | | | | NeighGO_term = cytoplasm_:
membrane_(2.75/1.34)
ribosome_biogenesis_and_assembly_: membrane (0.0)
| | \ | | - | | | \ | Neig}hGO_term =
mitochondrion_: membrane_(2.09/0.79)
| | | | | | | | | | NeighGO_term = RNA_binding_:
membrane_(0.0)
| | | | | | | | | | NeighGO_term =
RNA_metabolic_process_: membrane_ (0.0)
| | | | | | | | \ | NeighGO_term = membrane_:
membrane_(0.0)
| | | | | | | | | | NeighGO_term = ribosome_:
membrane_(0.0)
| | \ | | | | | | | NeighGO_term =
hydrolase_activity_: membrane_ (0.0)
| | | | | | | | | | NeighGO_term = nucleus_:
membrane_ (0.0)
protein_binding_: membrane_ (0.0)
| | | | | | | - | | | NeighGO_term =
protein_catabolic_process_: membrane_ (0.0)
| | | | | | | | | | NeighGO_term =
biological_process_: membrane_(2.45/0.93)
| | | | | | | | | NeighGO_term =
membrane_fraction_: membrane_ (0.0)
| | \ | | | | \ | | NeighGO_term =
signal_transducer_activity_: membrane_ (0.0)
| | | | | | | | | | NeighGO_term =
cellular_component_: membrane_ (0.0)
| | | | | | | | | NeighGO_term =
transcription_regulator_activity : membrane_ (0.0)
| | | | | | | | | | NeighGO_term =
DNA_metabolic_process_: membrane_(0.0)
| | | | | | | | | | NeighGO_term = translation_:
nucleus_(1.23/0.6)
| | | | | | | | | | NeighGO term = chromosome :
membrane_(0.0)
| | | | | | | | | | NeighGO_term = cell_cycle_:
membrane_(0.0)
| | | | | | | | | | NeighGO_term =
plasma_membrane_: membrane_(0.0)
| | | | - | | | - | | | NeighGO_term =
transferase_activity_: membrane_ (0.0)
| | | | | | | | | | NeighGO_term =
structural_molecule_activity_: membrane_ (0.0)
| | | | | | | | | | NeighGO_term = vesicle-
mediated_transport_: cellular_component_(8.06/\overline{2.49)}
| | | | | | | | | | - NeighGO_term =
transcription_: membrane_ (0.0)
| | | | | | | | | | NeighGO_term =
endoplasmic_reticulum_: membrane_ (0.0)
```

$|\quad| \quad|\quad| \quad \mid \quad$ NeighGO_term $=$ oxidoreductase_activity_: membrane_ (0.0) | | | | | | | | $\mid$ NeighGO_term = lipid_metabolic_process_: membrane_ (0.0)
 membrane (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGo_term $=$ cell_cortex_: nucleus_(4.83/2.41)
| | | | | | | | | NeighGO_term = cytoskeleton_: nucleus_(2.3/0.76)
 lyase_activity_: membrane_ (0.0)
| $\left.\right|^{-}$| $\left.\left.\right|^{-}|\quad|\right|^{-}|\quad| \quad$ NeighGO_term $=$ amino_acid_and_derivative_metabolic_process_: membrane_ (0.0) | | | | | | | | | NeighGO_term = ligase_activity_: membrane_ (0.0)
| | | | | | | | | | NeighGO_term = cell_wall_organization_and_biogenesis_: membranē (0.0) | $\left.\left|\left.\right|^{-}\right| \quad\left|\left.\right|^{-}\right|^{-}|\quad|\right|^{-}$NeighGO_ term = cellular_bud_: membrane_ (0.0)
 site_of_polarized_growth_: membrane_ (0.0)
$\mid \quad$ | | | $\mid \quad\rceil|\quad| \quad$ | NeighGO_term = motor_activity_: membrane_ (0.0)
| | | | | | | | | ${ }^{-}$NeighGO_term = cytokinesis_:
membrane_(0.0)
$|\quad| \quad|\quad| \quad \mid \quad$ NeighGo_term = peptidase_activity_: membrane_(0.0)
| | $\left.\right|^{-}|\quad| \quad|\quad| \quad \mid \quad$ NeighGo_term $=$ protein_modification_process_: membrane_(0.0)
| | | | | $\mid$ | $|\quad| \quad \mid \quad$ NeighGO_term = electron_transport_: membrane_ (0.0)
| | $\mid$ | | | | ${ }^{-}$| ${ }^{-}$NeighGOterm = carbohydrate_metabolic_process_: membrane (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NēighGo term $=$ enzyme_regulator_activity_: membrane_(0.0)
 cell_budding_: membrane_(0.0)
$|\quad| \quad$ | | | - | | | NeighGO term = nucleolus : membrane_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGO_term = conjugation_:
membrane_(0.0)
$|\quad| \quad|\quad| \quad \mid \quad$ NeighGO_term $=$ isomerase_activity_: membrane_(0.0)
| | | | | | | | ${ }^{-}$| NeighGo_term = phosphoprotein_phosphatase_activity_: membrane_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ | NeighGO_term = protein_kinase_activity_: membrane_ (0.0)
$\mid$ | | | | | NeighGO term = membrane_organization_and_biogenesis_: membrane_ (0.0) $|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGO_term $=$ Golgi_apparatus_: membrane_(0.0)
| | | | | | | | | NeighGO_term = cytoskeleton_organization_and_biogenesis_: membrane_ (0.0)


```
| | | | | | | | NeighGO_term = response_to_stress_:
site_of_polarized_growth_(0.0)
| | | | | | | | NeighGO_term = cytoplasmic_membrane-
bound_vesicle_: site_of_polarized_growth_(0.0)
transporter_activity_: site_of_polarized_growth_(0.0)
| | | | | \ | - | NeighGO_E
site_of_polarized_growth_(0.0)
| | | | | | | | NeighGO_term = cytoplasm_:
site_of_polarized_growth_ (0.0)
| | | | | | | | NeighGO_term =
ribosome_biogenesis_and_assembly_: site_of_polarized_growth_(0.0)
| | | | | - | | | N
cellular_bud_(1.41/0.53)
| | | | | | | | NeighGO_term = RNA_binding_:
site_of_polarized_growth_ (0.0)
RNA_metabolic_process_: site_of_polarizēd_growth_(0.0)
| | | | | | | | NeighGO_term = membrane_:
site_of_polarized_growth_(0.0)
| | | | | | \ | NeighGO_term = ribosome_:
site_of_polarized_growth_ (0.0)
| \ | | | | ` | NeighGO_term = hydrolase_activity_:
site_of_polarized_growth_(0.0)
| | | | | | | | NeighGO_term = nucleus_:
site_of_polarized_growth_ (0.0)
| \ | | | | \ | NeighGO_term = protein_binding_:
site_of_polarized_growth_(0.0)
| \ - | | | \ | NeighGO_term =
protein_catabolic_process_: site_of_polarized_growth_(0.0)
| | | | | | | NeighGO_term = biological_process_:
cellular_bud_(1.19/0.49)
| | | | | | | | NeighGO_term = membrane_fraction_:
site_of_polarized_growth_(0.0)
| | | | | | \ | NeighGO_term =
signal_transducer_activity_: site_of_polarized_growth_ (0.0)
| | | | | | | | NeighGO_term = cellular_component_:
cellular_bud_(8.29/3.1)
| | \ \ | | | | NeighGO_term =
transcription_regulator_activity_: site_of_polarized_growth_(0.0)
| | | | | | | | NeighGO_term =
DNA_metabolic_process_: site_of_polarized_growth_ (0.0)
| | | | | | | | NeighGO_term = translation_:
site_of_polarized_growth_ (0.0)
site_of_polarized_growth_(0.0)
| | | | | | | | NeighGO_term = cell_cycle_:
site_of_polarized_growth_ (0.0)
| \ | | | | ` | NeighGO_term = plasma_membrane_:
site_of_polarized_growth_ (0.0)
| | | | | | | NeighGO_term =
transferase_activity_: cellular_bud_ (3.03/1.01)
| | | | | | | | NeighGO_term =
structural_molecule_activity_: site_of_polarized_growth_ (0.0)
```

```
| | | | | | | | NeighGO term = vesicle-
mediated_transport_: site_of_polarized_growth_ (0.0)
| | | | | | | | NeighGO_term = transcription_:
site_of_polarized_growth_ (0.0)
endoplasmic_reticulum_: site_of_polarized_growth_ (0.0)
| | | | | | | | ` -NeighGO tērm =
oxidoreductase_activity_: site_of_polarized_growth_ (0.0)
| | | | | | | | NeighGO_term =
lipid_metabolic_process_: site_of_polarized_growth_(0.0)
| | | | | | | | NeighGO_term = meiosis_:
site_of_polarized_growth_(0.0)
| | | | | | | | NeighGO_term = cell_cortex_:
site_of_polarized_growth_(0.0)
| | | | | | | NeighGO_term = cytoskeleton_:
site_of_polarized_growth_(0.0) NeighGO term = lyase activity :
site of polarized growth (0.0)
| | | | | | | | NeighGO_term =
amino_acid_and_derivative_metabolic_process_: cellular_bud_
(6.94/1.79)
| | | | | | | NeighGO_term = ligase_activity_:
site_of_polarized_growth_(0.0)
| | | | | | | | NeighGO_term =
cell_wall_organization_and_biogenesis_: site_of_polarized_growth_
(5.26/1.06)
| | | | | | | | NeighGO_term = cellular_bud_:
site_of_polarized_growth_(0.0)
| | | | | | \ | NeighGO_term =
site_of_polarized_growth_: site_of_polarized_growth_(0.0)
| | | | | | | NeighGO_term = motor_activity_:
site_of_polarized_growth_(0.0)
| | | | | | \ | NeighGO_term = cytokinesis_:
site_of_polarized_growth_(0.0)
| | | | | | \ | NeighGO_term = peptidase_activity_:
cellular_bud_(1.1/0.21)
| | | | | | | | NeighGO_term =
protein_modification_process_: site_of_polarized_growth_(3.4/0.8)
| | | | | \ | | NeighGO_term = electron_transport_:
site_of_polarized_growth_(0.0)
| | | | | | | | NeighGO_term =
carbohydrate_metabolic_process_: site_of_polarized_growth_(0.0)
| | | | | | | | NeighGO_term =
enzyme_regulator_activity_: site_of_polarized_growth_(0.0)
```



```
site_of_polarized_growth_(0.0)
| \ | | | | ` | NeighGO_term = nucleolus_:
site_of_polarized_growth_ (0.0)
| \ | | | | ` | NeighGO_term = conjugation_:
site_of_polarized_growth_(0.0)
| | | | | | | | NeighGO_term = isomerase_activity_:
site_of_polarized_growth_ (0.0)
```

```
| | | | | | | | NeighGO term =
phosphoprotein_phosphatase_activity_: sīte_of_polarized_growth_
(0.0)
| | | | | | | | NeighGO_term =
protein_kinase_activity_: site_of_polarized_growth_(0.0)
membrane organization and biogenesis : site of polarized growth
(0.0)
| | | | | | | NeighGO_term = Golgi_apparatus_:
site_of_polarized_growth_ (0.0)
| | | | | | \ | NeighGO_term =
cytoskeleton_organization_and_biogenesi\overline{s}_: site_of_polarized_growth
(0.0)
| | | | | | | | NeighGO_term =
cellular_respiration_: site_of_polarized_growth_(0.0)
| | | | | | | | NeighGO_\overline{term = DNA_binding_:}
site_of_polarized_growth_(0.0)
| | | | | | \ | NeighGO_term = sporulation_:
site_of_polarized_growth_(0.0)
| \ | | | | \ | NeighGO_term = peroxisome_:
site_of_polarized_growth_ (0.0)
| \ - | | | ` | NeighGO_term =
organelle_organization_and_biogenesis_: site_of_polarized_growth_
(0.0)
| | | | | | | | NeighGO_term =
anatomical_structure_morphogenesis_: site_of_polarized_growth_ (0.0)
| | | | | \ | | Ne\overline{ighGO_t\overline{erm = helicase_activi}ty_:}
site_of_polarized_growth_ (0.0)
| \ | | | | \ | NeighGO_term = cell_wall_:
site_of_polarized_growth_(0.0)
| | | | | | | | NeighGO_term = vacuole_:
site_of_polarized_growth_(0.0)
| \ | | | | ` | NeighGO_term = pseudohyphal_growth_:
site_of_polarized_growth_(0.0)
| | | | | | \ | NeighGO_term =
nuclear_organization_and_biogenesis_: site_of_polarized_growth_
(0.0)
| | | | | | | | NeighGO_term = signal_transduction_:
site_of_polarized_growth_ (0.0)
| | | | | | | | NeighGO_term =
translation_regulator_activity_: site_of_polarized_growth_(0.0)
| | | | | | | | NeighGO_term =
generation_of_precursor_metabolites_and_energy_:
site_of_polarized_growth}_(0.0
| | | | | | \ | NeighGO_term =
vitamin_metabolic_process_: site_of_polarized_growth_ (0.0)
| | | | | | | | NeighGO_term =
extracellular_region_: site_of_polarized_growth_(0.0)
| | | | | \ | | | NeighGO_Eerm =
nucleotidyltransferase_activity_: site_\overline{of_polarized_growth_(0.0)}
| | | | | | | | NeighGO_term =
mitochondrial_envelope_: site_of_polarized_growth_ (0.0)
| | | | | | | | NeighGO_term =
microtubule_organizing_center_: site_of_polarized_growth_ (0.0)
```











| | | | | | | $\mid$ NeighGO_term $=$ mitochondrial_envelope_: cytoplasm_(0.0) $|||||||\mid$ NeighGO_term $=$ microtubule_organizing_center_: cytoplasm_(0.0)
$|||||||\quad| \quad| \quad$ start $>$ 352759: cytoplasm_ (35.99/15.42)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>353628$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid \quad$ start $<=358572$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid \quad$ start $<=354818:$ |  |  | endoplasmic_reticulum_(16.08)

| | | | | $\left.\right|^{-}|\quad| \quad|\quad| \quad$ start $>$ 354818: chromosome_ (9.85)
 (12.37)

cytoplasm_(29.92/10.85)
$\mid$ | | | | | | | | $\mid$ start > 370057: cellular_component_ (18.67)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |

nucleus_(16.0/4.06)
$\mid$ | | | | | | | $\mid$ | $\mid$ neigh_strand $=C$ : cytoplasm_(7.53/2.51)
| | | | start > 377702
| | | | | start $<=412910$
| | | | | | start $<=385513$
| | | | | | | tart $<=381543$

(12.94)
| | | | | | | | | | start > 388729
| | | | | | | | | | start <= 390303:
nucleus_(11.74)
| | - | | | | | | | | start > 390303: nucleolus_ (7.27)
| | | | | | | | | start > 393537: peroxisome_ (14.8)

| | | | | | | | | | distance <= 1819: endomembrane_system_(4.88/1.44) $|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ distance $>$ 1819: membrane_(5.48/2.74)
$|||\quad| \quad| \quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ Neigh_GO_aspect $=P:$ endoplasmic_reticulum_(11.35/5.71)
| | | - | | $\mid$ | | start > 454229:
cellular_component_(17.48)

nucleus_(11.09)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>459299$ |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | neigh_num $<=1$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | distance $<=$ |

1809: mitochondrion_ $\underset{\mid}{(5.46 / 2.19)}$ | | | | | | | | | | | |

1809: nucleus_(5.9/2.21)
 cytoplasm (19.2/9.06)
$\mid$
$\mid$
$\mid$
$\mid$
$\mid$
$\mid$
$\mid$
$\mid$
$\mid$$|$
| | | | | | | | | | | | start > 467228: cytoplasm_(12.38)

membrane_(19.3)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ start $>475778$
| | | | | | | | | | start <= 480990:
cytoplasm_(39.0/19.36)
| | | | | | | | | | | | start > 480990:
mitochondrion_(15.11)
$\begin{array}{llllllll}\mid & \mid & \mid & \mid & \mid & \text { start }>484028 \\ \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=484845 \text { : }\end{array}$
endoplasmic_reticulum_(19.29)
| | | | | | | | | | start > 484845: cytoplasm_
(21.41)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>486826$ |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=499079$ |

$|||||||||\mid$ start<= 488657:
nucleus_(19.51)

```
| | | | | | | | | | | | start > 488657:
endomembrane_system_(15.05)
|
nucleolus_(21.8/8.17)
| | | | | | | | | | | | start > 491931:
nucleus_(17.63)
| | | | | | | | | | start > 493896:
cellular_component_ (12.34)
| |lllllllllll
vacuole_(16.85)
| | | | | | | | | | | start > 502388:
endoplasmic_reticulum_(13.57)
```



```
nucleus_(47.59/9.86)
| | | | | | | | | | | start > 512732:
mitochondrion_(19.57)
| start > 5\overline{1}7532
| | start <= 551499
| | | start <= 525392
| | | | start <= 521737: cellular_component_(45.05)
| | | | start > 521737: cell_wall_ (14.22)
| start > 525392: cellular_component_ (98.67)
start > 551499
start <= 554396
| start <= 552099: vacuole_ (9.41)
| start > 552099: cytoplasm_ (9.9)
start > 554396: cellular_component_ (18.2)
```


## For Molecular Function

```
| start <= 209699
| | start <= 151299
| | | start <= 117809
| | | | start <= 85367
| | | | | start <= 34075
| | | | | | start <= 21780
| | | | | | | start <= 3310: helicase_activity_
                            (5.29)
| | | | | | | | | | | start > 3310 
| | | | | | | | | start <= 15665
| | | | | | | | | | start <= 6400:
```



```
| | | | | | | | | | start<= 10211:
transporter_activity_ (6.44)
| | | | | | | | | | | start > 10211:
molecular_function_(11.25)
| | | | | | | | | start > 15665:
transporter_activity_ (12.29)
```

```
(60.12)
```



```
(10.47)
| | | | | | strand = W: transporter_activity_
(12.69)
| | | | | start > 34075
| | | | | | start <= 40082
| | | | | | | | start <= | | start <= 36023:
structural_molecule_activity_ (13.19)
| | | | | - | | \ start > 36023: transferase_activity_
(13.64)
| | | | | | start > 38506: protein_binding_ (25.43)
| | | | | start > 40082
| | | | | | start <= 69545
| | | | | | | | | | | | | in start <= 67453 
| | | | | | | | | | start <= 65856
| | | | | | | | | | | start<= 62753:
molecular_function_(63.35) | | | | start > 62753
| | | | | | | | | | | | | | | start <= 64155:
hydrolase_activity_ (10.34)
| | | | | | | | | | | | start > 64155:
molecular_function_(9.42)
| | | | | | | | | | start > 65856:
transcription_regulator_activity_ (10.52)
| | | | | | | | | strand = W
```



```
signal_transducer_activity_ (14.45)
| | | | |
transcription_regulator_activity_ (45.23)
| | | | | | | | | | start > 54849
| | | | | | | | | | | start<= 56647:
RNA_binding_(8.14)
| | | | | | | | | | | start > 56647:
molecular_function_(11.13)
| | | | | | | | start > 67453: protein_binding_
(19.05)
\begin{tabular}{lllllllll}
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & start \(>69545\) \\
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & start \(<=78932\) \\
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & start \(<=74241\) \\
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid \quad\) strand \(=\mathrm{C}:\) \\
transporter activity & \((11.22)\)
\end{tabular}
| | | | | | | | | | strand = W:
molecular_function_(28.66)
```












```
| | | | | | | | | | start > 480990:
hydrolase_activity_ (12.22)
| | | | | | | | | start > 484028
| | | | | | | | | | start<= 484845:
transferase_activity_ (21.12) | | start > 484845:
molecular function_(32.44)
```



```
protein_binding_ (17.68)
| | | | | | | | | start > 490747: RNA_binding_
(19.31)
| | | | start > 491931
| | | | | start <= 512732
| | | | | | | | start <= 506319 
| start <= 493896: molecular function
(17.91)
| | | | | | | | start > 493896: peptidase_activity_
(14.64)
| | | | | | start > 499079: molecular_function_
(37.22)
| | | | | | start > 506319
| | | | | | | start <= 509363:
protein_kinase_activity_ (22.34)
| | | | | | | start > 509363: DNA_binding_
(48.3/30.44)
| | | | | start > 512732
| | | | | start <= 521737: transferase_activity_
(38.99)
| | | | | | start > 521737
| | | | | | start <= 541651: molecular_function_
(78.06)
| | | | | | | start > 541651
| | | | | | | start <= 554396
| | | | | | | | | start <= 552099:
hydrolase_activity_ (10.53)
| | | | | | | | | start > 552099:
oxidoreductase_activity_ (6.69)
| | | | | | | | start > 554396: molecular_function_
(19.14)
```


## For Biological Process

```
start <= 452872
| | start <= 219886
| | | stop <= 149568
| | | | start <= 48761
| | | | | start <= 34075
| | | | | | start <= 15665
| | | | | | | start <= 10211
| | | | | | | strand = C
biological_process_ (22.31)
```




```
| | | | | | | | | | | | neigh num <= 1:
biological_process_ (18.12/6.17)
| | | | | | | | | | | | | neigh_num > 1:
response_to_stress_ (13.22/2.93)
| | | | | | | | | | | | start > 123585:
biological_process_ (16.92)
| | | | | | | | | | start > 127774:
DNA_metabolic_process_ (15.0)
|
translation_(11.54)
| | | | | | | | | | start > 134547: vesicle-
mediated_transport_ (11.91)
```



```
biological_process_ (13.51/3.39)
| | | | | | | | | | | neigh_num > 1:
amino_acid_and_derivative_metabolic_process_ (13.65/5.75)
| | | | | | | | | | start > 141395
| | | | | | | | | | | start <= 149218:
translation (30.66)
| | | - | | | | | | | start > 149218:
biological_process_ (19.18)
| | | | | strand = W
| | | | | | start <= 56647
| | | | | | start <= 54849
| | | | | | | | start <= 51109: meiosis_(18.25)
start > 51109:
DNA_metabolic_process_ (11.57)
| | | | | | | | | | | start > 54849 
sporulation_(0.0)
| | - | | | | NeighGO_term = molecular_function_:
meiosis_ (9.58/2.78)
sporulation_(0.0)
| | | | | | | | NeighGO_term = cytoplasmic_membrane-
bound_vesicle_: sporulation_(0.0)
| | | | | | | | NeighGO_term =
transporter_activity_: sporulation_ (0.0)
| | | | | | | | NeighGO_term = transport_:
sporulation_(0.0)
| | | | | | | | NeighGO_term = cytoplasm_:
sporulation_(1.9/0.5)
| | | | | | | | NeighGO_term =
ribosome_biogenesis_and_assembly_: sporulation_ (0.0)
| | | | | | | | NeighGO_term = mitochondrion_:
sporulation_(0.0)
| | | | | | | | NeighGO_term = RNA_binding_:
sporulation_(0.0)
| | | | | | | | NeighGO_term =
RNA_metabolic_process_: sporulation_(0.0)
```



```
| | | | | | | | NeighGO term =
amino_acid_and_derivative_metabolic_proc
| | | | | | | | Neig}hGO_term = ligase_activity_:
sporulation_(0.0)
| | | | | | | | NeighGO_term =
cell_wall_organization_and_biogenesis_: sporulation_(0.0)
| ` | | | | - | - | Neigh\overline{GO_term = cellūlar_bud_:}
sporulation_(0.0)
| | | | | | | | NeighGO_term =
site_of_polarized_growth_: sporulation_ (0.0)
| \ | | | | ` | NeighGO_term = motor_activity_:
sporulation_(0.0)
| | | | | | | | NeighGO_term = cytokinesis_:
sporulation_(0.0)
| | | | | | | NeighGO_term = peptidase_activity_:
sporulation_(0.0)
| | | | | | | | NeighGO term =
protein_modification_process_: sporulation_(0.0)
| | | | | \ | \ NeighGO_term = electron_transport_:
sporulation_(0.0)
| | | | | | | | NeighGO_term =
carbohydrate_metabolic_process_: sporulātion_ (0.0)
| | | | | | | | - NeighGO_term =
enzyme_regulator_activity_: sporulation_(0.0)
| | | | | | | | NeighGO_term = cell_budding_:
sporulation_(0.0)
| | | | | | | | NeighGO_term = nucleolus_:
sporulation_(0.0)
| | | | | | | | NeighGO_term = conjugation_:
sporulation_(0.0)
| | | | | | | | NeighGO_term = isomerase_activity_:
sporulation_(0.0)
| | | | | | | | NeighGO_term =
phosphoprotein_phosphatase_activity_: s\overline{porulation_(0.0)}
| | | | | | | | NeighGO_term =
protein_kinase_activity_: sporulation_(0.0)
| | | | | | | | NeighGO_term =
membrane_organization_and_biogenesis_: sporulation_(0.0)
| | \ | | | | | | Neig\overline{hGO_term = Golg}\mp@subsup{]}{_}{-}apparatus_:
sporulation_(0.0)
| | | | | | | | NeighGO_term =
cytoskeleton_organization_and_biogenesis_: sporulation_(0.0)
| | | | | | | | NeighGO_term =
cellular_respiration_: sporulation_(0.\overline{0})
| | \ | \ | | NeighGO_term = DNA_binding_:
sporulation_(0.0)
| | | | | | | | NeighGO_term = sporulation_:
sporulation_(0.0)
| | | - | | | | NeighGO_term = peroxisome_:
sporulation_(0.0)
| | | | | | | | NeighGO_term =
organelle_organization_and_biogenesis_: sporulation_ (0.0)
| | | | | | | | NeighGO_term =
anatomical_structure_morphogenesis_: sporulation__ (0.0)
```




```
| | | | | | | | | start <= 159431:
amino_acid_and_derivative_metabolic_process_ (34.68)
| | | | | | | | | -start > 159431:
organelle_organization_and_biogenesis_ (18.4)
| | | | | | start > 167427
| | | | | | | strand = C
| | | | | | | | | | | | in start <= 170337 
biological_process_ (17.32)
| | | | | | | | | start > 168554:
cell_wall_organization_and_biogenesis_(22.12/9.39)
| \ | | | | | - | - | start 
| | | | | | | | | start <= 173337:
DNA_metabolic_process_ (18.8)
| | | | | | | | | start > 173337:
ribosome_biogenesis_and_assembly_ (19.63)
| | \ | | - | | strañd = W: biological_process_ (51.53)
| | | | | start > 178212
| | | | | | strand = C
| | | | | | | start <= 186802: DNA_metabolic_process_
(49.99/29.93)
| | | | | | | start > 186802:
organelle_organization_and_biogenesis_ (17.45)
| | | | | | - střand = W
| | | | | | | start <= 180338
| | | | | | | | neigh_num <= 1:
nuclear_organization_and_biogenesis_(11.71/3.14)
| | | | | \ \ | neig}h_num > 1:
membrane_organization_and_biogenesis_-(17.35/4.82)
| | | | | | | | | | start > 180338 
amino_acid_and_derivative_metabolic_process_(23.08)
| | | | | | | | | start > 181\overline{970: translation_ (15.96)}
| | | | start <= 194801
| | | | strand = C
| | | | | | | start <= 190536: DNA_metabolic_process_
(20.24/9.2)
| | | | | | | start > 190536
| | | | | | | | start <= 193538: response_to_stress_
(14.36)
| | | | | | | | start > 193538:
carbohydrate_metabolic_process_ (14.59)
| | | \ | | - strand = W
| | | | | | | start <= 187917: RNA_metabolic_process_
(13.46)
| | | | | | | start > 187917: electron_transport_
(17.37)
| | | | | start > 194801
| | | | start <= 210842
| | | | | | | strand = C
| | | | | | | | start <= 198278:
carbohydrate_metabolic_process_ (7.29)
```





```
| | | | | | | | | | NeighGO term =
electron_transport_: DNA_metabolic_process_ (0.\overline{0})
| | | | | | \ | | | | NeighGO_term =
carbohydrate_metabolic_process_: DNA_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO_term =
enzyme_regulator_activity_: DNA_metabolic_process_ (0.0)
| | | | \ | | | - | | NēighGO_tērm =
cell_budding_: DNA_metabolic_process_ (0.0)
| | | | | | | | | NeighGO_term = nucleolus_:
DNA_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO_term = conjugation_:
DNA_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO_term =
isomerase_activity_: DNA_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO_term =
phosphoprotein_phosphatase_activity_: DNA_metabolic_process_(0.0)
| | | | | | | | | | NeighGO term =
protein_kinase_activity_: DNA_metabolic_process_ (1.72/0.58)
| | | | | | - | | | | - NeighGO_term =
membrane_organization_and_biogenesis_: DNA_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO_term =
Golgi_apparatus_: DNA_metabolic_process_ (0.0)
| |
cytoskeleton_organization_and_biogenesis_: DNA_metabolic_process_
(0.0)
| | | | | | | | | | NeighGO_term =
cellular_respiration_: DNA_metabolic_process_(\overline{0}.0)
| | \ | | \ | | | | | Neigh̄GO_term = DNA_binding_:
DNA_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO_term = sporulation_:
DNA_metabolic_process_(0.0) | | NeighGO term = peroxisome :
DNA_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO_term =
organelle_organization_and_biogenesis_: DNA_metabolic_process_(0.0)
| | | | | | | | | | NeighGO_term =
anatomical_structure_morphogenesis_: DNA_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO_term =
helicase_activity_: DNA_metabolic_process_ (0.0)
| | | | | | | | | | | NēighGO_term = cell_wall_:
DNA_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO_term = vacuole_:
DNA_metabolic_process_ (0.0)
| | | | | | | | | NeighGO_term =
pseudohyphal_growth_: DNA_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO_term =
nuclear_organization_and_biogenesis_: DNA_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO_term =
signal_transduction_: DNA_metabolic_process_ (0.0)
| | | | | - | |
translation_regulator_activity_: DNA_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO_term =
generation_of_precursor_metabolites_and_energy_:
DNA_metabolic_process_ (0.0)
```



```
| | | | | | | | | start <= 242584:
pseudohyphal_growth_(12.37)
| | | | | | | | | | start > 242584:
vitamin_metabolic_process_ (7.92)
| | | | | | | | | start > 246195:
biological_process_(29.2)
| | | | | | | | | start > 261593
| | | | | | | | | start <= 267540:
DNA_metabolic_process_ (18.12)
| | | | | | | | | start > 267540
| | | | | | | | | start <= 272629:
membrane organization and biogenesis (22.07)
| | | | | | | | | \ start > 272629:
pseudohyphal_growth_(17.79)
| | | | | | | start > 274176
| | | | | | | | | | start <= 280233 
ribosome_biogenesis_and_assembly (15.92)
| | | | | | - | | \ start > 276766:
biological_process_ (35.38)
| | | | | | | | start > 280233
| | | | | | | | | start <= 280822:
RNA metabolic process (15.31)
| | | | | |- | | | start > 280822:
ribosome_biogenesis_and_assembly_ (19.47)
| | | | start > 286772
| | | | | start <= 328306
| | | | | start <= 314676
| | | | | | strand = C
| | | | | | | | start <= 296450: transport_ (44.29)
| | | | | | | | start > 296450
| | | | | | | | | start <= 302764
| start <= 298612:
biological process (12.77)
| | | | | | | | | | start > 298612
| | | | | | | | | | neigh_num<= 1:
transport_ (13.1/5.5)
| | | | | | | | | | | neigh_num > 1: vesicle-
mediated_transport_ (8.6/3.34)
| | | | | | | | | start > 302764:
biological_process_ (14.23)
| | | | | | | strand = W: DNA_metabolic_process_
(25.45)
| | | | | | start > 314676
| | | | | | start <= 328039
| | | | | | | | start <= 320417:
cell_wall_organization_and_biogenesis_ (39.33)
| | | | | | | | start > 320417
| | | | | | | | | strand = C:
cell wall organization and biogenesis (30.96/11.22)
| | | | | | | | | strand = W
| | | | | | | | | | start <= 323412:
response_to_stress_ (22.58)
| | | - | - | | | | | start > 323412
```








## Decision Tree Generated for Chromosome Nine:

## For Cellular Component

start $<=276524$
$\mid$
$\mid$














## For Molecular Function












## For Biological Process

start $<=128151$
$\mid$
$\mid$
$\mid$



| | | | | | | | | | start > 180424:
ribosome_biogenesis_and_assembly_ (16.06)
$\mid$
$\mid$
$\mid$

$\mid$$|$|  | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>183124$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| translation_(18.51) |  |  |  |  |  |  |  |

$|\quad| \quad$ | | | | | | | | start > 183934:
amino_acid_and_derivative_metabolic_process_ (19.35)
| $\left.\left.\left.\left.\right|^{-}\right|^{-}\right|^{-}|\quad|\right|^{-}|\quad| \quad{ }^{-} \mid \quad$ start > 187629: translation_ (22.95)

biological_process_(12.97)
| | | | | | | | | | | start > 200116 | | | | | | | | | | start <= 202040: cell_wall_organization_and_biogenesis_ (11.68)
| $\mid$ | | | | ${ }^{-}$| | | ${ }^{-}$| start > 202040:
protein_modification_process_ (10.74)
| | $\mid$ | $|\quad| \quad$ | | | start > 203256
| | | | | | | | | | start <= 204650:
biological_process_ (16.31)

$|$| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>204650$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=212496$ : |  |

RNA_metabolic_process_ (19.09)
| ${ }^{-} \mid$| ${ }^{-}$| | | | | |
biological_process_ (16.23)
| | | | | | | | start > 216655
| | | | | | | | | | start <= 222487
| | | | | | | | | | | start<= 220697:
protein_catabolic_process_ (19.02) | | | | | |
amino acid and derivative metabolic process (33.3)
| | | | | | | | | start > 222487

protein_modification_process_ (5.11)
| | - | | $\quad \mid \quad \overline{s t a r t ~>~} 229991$
| | | | | | | | start <= 235471
$\begin{array}{lllllll}\mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=232366 \\ \mid & \mid & \mid & \mid & |\quad| \quad \text { start }<=231069:\end{array}$
cellular_respiration (9.82)
| | $\mid$ | | | | | | start > 231069: translation_ (10.47)
(13.14)

| | | | | | | | start <= 300008: protein_catabolic_process_(8.75)
| | - | | | | | | $\left.\right|^{-}$start > 300008: translation_ (17.62)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>300298$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=304102:$ |

biological_process (33.58)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>304102$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| start $<=307927:$ |  |  |  |  |  |  |  |  |

cell_wall_organization_and_biogenesis_ (13.97)
| | | | | | | | | | start > 307927:
biological_process_ (13.53)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>309169$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=322340$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ start $<=316766$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ start $<=311163:$ |  |

transport_ (14.75)
| | | | | | | | | | start > 311163:
amino_acid_and_derivative_metabolic_process_ (20.73)
| $\left.\left.\left.\left.\right|^{-}\right|^{-}\right|^{-}|\quad|\right|^{-}|\quad| \quad$ start $>$ 316766: sporulation_
(38.64/17.18)

biological_process_ (21.61)
| | | | | ${ }^{-}$| | | | | start > 326101
| | | | | | | | | | | start <= 333724:
transport_ (12.43)
| | | | | | | | | | | | start > 333724
| | | | | | | | | | | | | start <= 336209:
DNA metabolic_process (11.79)
| | | | | | | | | | | | | start > 336209:
translation_(5.98)
| | | | | | | | | start > 339341:
protein_catabolic_process_ (21.49)
| | | | | | |- start > 345689
| | | | | | | start $<=353937$
$\begin{array}{lllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=350298 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { neigh_strand }=W \text { : vesicle- }\end{array}$
mediated_transport_(17.05/7.53)
| | $\mid$ | ${ }^{-}|\quad| \quad|\quad| \quad$ neigh_strand = C: transport_
(15.93/5.23)
| | | | | | | | | start > 350298:
cell_wall_organization_and_biogenesis_ (26.91)
| | | | | | | | start > 353937
| | | | | | | | start <= 356892:
RNA_metabolic_process_(14.99)
| ${ }^{-}$| | | | $\left.\right|^{-}$| | start > 356892
| | | | | | | | | $\mid$ start <= 364886:
DNA_metabolic_process_(11.42)






## Decision Tree Generated for Chromosome Ten:

 For Cellular Component

```
| | | | | | | | | NeighGO_term = transport_:
nucleus (0.0)
| | | | | | | | | NeighGO_term =
endoplasmic_reticulum_: nucleus_ (0.0)
| | | | | | | | | NeighGO_term = peroxisome_:
nucleus_ (0.0)
| | | | | | | | | NeighGO_term = protein_binding_:
nucleus (0.0)
| | | | | | | | | NeighGO_term =
organelle_organization_and_biogenesis_: nucleus_ (0.0)
| | | | | | | | | NeighGO_term = mitochondrion_:
mitochondrion (2.43/0.21)
| | | | | | | | | NeighGO_term = nucleus_:
nucleus_(6.04/1.66)
| | | | | | | | | NeighGO_term =
ribosome_biogenesis_and_assembly_: nucleus_(0.0)
| | | | | | | | \ NeighGO_term = RNA_binding_:
nucleus (5.99/0.46)
| | | | | | | | | NeighGO_term =
RNA_metabolic_process_: mitochondrion_(3.04/0.56)
| | | | | | | | NeighGO_term = lyase_activity_:
nucleus (0.0)
| | | | | | | | | NeighGO term =
DNA_metabolic_process_: nucleus_ (0.0)
| | | | | | | | | NeighGO_term = cell_cycle_:
nucleus_ (0.0)
| | | | | | | | | NeighGO_term =
molecular_function_: nucleus_(1.4/0.18)
| | | | | | | \ | NeighGO_term = vesicle-
mediated_transport_: nucleus_ (5.76/1.34)
| | | | | | | | | NeighGO_term = nucleolus_:
nucleus_ (0.0)
| | | | | | | | | NeighGO_term = ribosome_:
mitochondrion (3.2/0.33)
| | | | | | | | NeighGO_term = translation_:
nucleus_ (0.0)
| | | | | | | | | NeighGO_term =
structural_molecule_activity_: nucleus_ (0.0)
| | | | | - | | | | NeíighGO_term = Golgi_apparatus_:
nucleus_(0.0)
| | | | | | | | | NeighGO_term =
transferase_activity_: nucleus_ (0.0)
| | | | | | | | | NeighGO_term =
protein_modification_process_: nucleus_ (0.\overline{0})
| | | | | \ | \ | NeighGO_term = chromosome_:
nucleus_(0.0)
| | | | | | | | | NeighGO_term = DNA_binding_:
nucleus_ (0.0)
| | | | | | | | | NeighGO_term =
response_to_stress_: nucleus_ (0.0)
| | \ | | | | \ | NeighGO_term = cytoplasm_:
nucleus_ (0.0)
| | | | | | | | | NeighGO_term =
cytoskeleton_organization_and_biogenesis_: nucleus_ (0.0)
```

```
| | | | | | | | | NeighGO_term = membrane_:
nucleus_(0.0)
| | | | | | | | | NeighGO_term =
enzyme_regulator_activity_: nucleus_ (0.0)
| | | | | | | | | NeighGO_term =
transcription_regulator activity_ nucleus_(0.0)
| | | | | - | | ` NeighGŌ_term =
extracellular_region_: nucleus_ (0.0)
| | | | | | | | | NeighGO_term =
cell_wall_organization_and_biogenesis_: nucleus_ (0.0)
| | | | | | | | | NeighGO_term = vacuole_:
nucleus_(0.0)
| | | | | | | | NeighGO_term =
peptidase_activity_: nucleus_ (0.0)
| | | | | | | | | NeighGO_term =
lipid_metabolic_process_: nucleus_ (0.0)
| | | | | | | | | NeighGO_term =
protein_kinase_activity_: nucleus_ (0.0)
| | | | - | | | | | | NeighGO_term =
cellular_homeostasis_: nucleus_ (0.0)
| | | | | | | | NeighGO_term = cell_wall_:
nucleus_ (0.0)
| | | | | | | | | NeighGO term =
nucleotidyltransferase_activity_ nucleus (
| | | | | | | | | NeighGO_term = meiosis_:
nucleus_ (0.0)
| | | | | | | | | NeighGO_term = transcription_:
nucleus_(0.0)
| | | | | | | | | NeighGO_term =
isomerase_activity_: nucleus_ (0.0)
| | | | | | | | | NeighGO_term =
carbohydrate_metabolic_process_: nucleus_(\overline{0}.0)
| | | | | | | | | NeighGO_term =
vitamin_metabolic_process_: nucleus_ (0.0)
| | | | | | | | | NeighGO_term =
translation_regulator_activity_: nucleus_ (0.0)
| | | | | | | | | NeighGO_term = ligase_activity_:
nucleus_ (0.0)
| | | | | | | | | NeighGO_term =
membrane_fraction_: nucleus_ (0.0)
| | | | | | | | | NeighGO_term =
helicase_activity_: nucleus_ (0.0)
| | | | | | | | | NeighGO_term = cell_cortex_:
nucleus_(0.0)
| | | | | | | | | NeighGO_term = cytokinesis_:
nucleus_(0.0)
| | | | | | | | | NeighGO_term =
nuclear_organization_and_biogenesis_: nucleus_(0.0)
| | | | | | | | | - NeighGO_tērm = conjugation_:
nucleus_(0.0)
| | | | | | | | | NeighGO_term =
oxidoreductase_activity_: nucleus_ (0.0)
| | | | | | | | NeighGO_term = sporulation_:
nucleus_ (0.0)
```







| | | | | | | | | | | NeighGO_term = peroxisome_: extracellular_region_ (0.0)
| | | | | | | | | | | protein_binding_: extracellular_region_ (0.0) organelle_organization_and_biogenesis_: extracellular_région_(0.0) | | ${ }^{-}$| | $\left.\left.\left.\right|^{-}\right|^{-}|\quad|\right|^{-}|\quad| \quad$ Neigh $\bar{G} O$ term ${ }^{-}=$ mitochondrion_: extracellular_region_(0.0) $|\quad| \quad|\quad| \quad|\quad| \quad$ | $\mid$ NeighGo_term $=$ nucleus_: plasma_membrane_(3.34/1.33)
 ribosome_biogenesis_and_assembly_: extracellular_region_ (0.0) | | $\mid$ | | | | | $\mid$ | | RNA_binding_: extracellular_region_ (0.0)
| | | | | | | | | | | 1 NeighGO_term = RNA_metabolic_process_: extracellular_region_ (0.0)
| | | | | | | | | | | ${ }^{-}$NeighGo_term = lyase_activity_: extracellular_region_ (0.0)
| | $\left.{ }^{-}\right|^{-}|\quad| \quad| |^{-}| |^{-}|\quad| \quad$ NeighGo_term $=$ DNA_metabolic_process_: extracellular_region_ (0.0)
 cell_cycle_: extracellular_region_ (0.0)
$\left.|\quad|{ }^{-}|\quad| \quad\left|{ }^{-}\right|\right|^{-}|\quad| \quad \mid \quad$ NeighGo_term $=$ molecular_function_: extracellular_region_(5.76/1.61)
 vesicle-mediated_transport_: extracellular_region_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ | $\mid$ NeighGo_term $=$ nucleolus_: extracellular_region_(0.0)
| | ${ }^{-}$| | | | $\quad$ | $\mid$ | NeighGo_term $=$ ribosome : extracellular region (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad$ | $\mid$ NeighGo_term $=$ translation_: extracellular_region_ (0.0)
| | | | | | | | | | | NeighGO term = structural_molecule_activity_: plasma_membrane_(2.33/0.85)
| | | $\quad$ | $\mid$ | $\left.|\quad|\right|^{-}|\quad|^{-}$NeighGo_term $=$ Golgi_apparatus_: extracellular_region_(0.0) $\quad$ (0) transferase_activity_: extracellular_region_ (0.0)
$|\quad| \quad-|\quad| \quad|\quad| \quad \mid \quad$ | NeighGoterm $=$ protein_modification_process_: extracellular_region_(0.0)
| | | | | | | | | | | | NeighGO_term = chromosome_: extracellular_region_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGo_term $=$ DNA_binding_: extracellular_region_ (0.0)
| ${ }^{-}\left|{ }^{-}\right| \quad|\quad|{ }^{-}| |^{-}|\quad| \quad$ NeighGO_term $=$ response_to_stress_: extracellular_region_ (0.0)
 cytoplasm_: endoplasmic_reticulum_(2.52/1.54)
| | | | | | | | | | | NeighGO_term = cytoskeleton organization and biogenesis : extracellular region (0.0)
| | | | | | | | | | | NeighGo_term = membrane_: endoplasmic_reticulum_(8.13/4.83)
| | | | | | | | | | NeighGO term = enzyme_regulator_activity_: extracellular_region_(3.52/1.09) $|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGo_term $=$ transcription_regulator_activity_: extracellular_region_(0.0) | | | | | | | | | | $\mid$ NeighGO_term = extracellular_region_: extracellular_region_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ | $|\quad| \quad$ NeighGo term $=$ cell_wall_organization_and_biogenesis_: extracellular_region_ (3.8 $\overline{3} / 1.1 \overline{5}$ )
 vacuole_: extracellular_region_ (0.0) $|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGo_term $=$ peptidase_activity_: extracellular_region_ (0.0)
| | | | | | | | | | | | NeighGo_term = lipid_metabolic_process_: extracellular_region_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGO_term $=$ protein_kinase_activity_: extracellular_region_ (0.0)
| | $|\quad|-|\quad| \quad|\quad| \quad{ }^{-} \mid \quad{ }^{-}$NeighGo_term = cellular_homeostasis_: extracellular_region_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad$ | $\mid \quad$ | cell_wall_: endoplasmic_reticulum_(8.11/4.59)
| | | | | | | | | | | NeighGo term = nucleotidyltransferase_activity_: extracellular_region_-(0.0)
| | | | | | | | | | NeighGO_term = meiosis_: extracellular region_(0.0)
| | | | | | | | | | | | NeighGo_term = transcription_: extracellular_region_(0.0)
| | | | | | | | | $\mid$ | $\quad$ NeighGo_term $=$ isomerase_activity_: extracellular_region_ (0.0)
| | | | | | | | | | | | NeighGO_term = carbohydrate_metabolic_process_: extracellular_region_(0.0)
 vitamin metabolic_process_: extracellular_region (0.0)
$|\quad| \quad{ }^{-}|\quad| \quad|\quad|-\quad|\quad| \quad \mid \quad \bar{N} e i g h G O$ term $=$ translation_regulator_activity_ extracellular_region $\overline{(0.0)}$
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGo_term $=$ ligase_activity_: extracellular_region_ (0.0)
 membrane_fraction_: extracellular_region_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGo_term $=$ helicase_activity_: extracellular_region_ (0.0)
| | | | | | | | | | | | NeighGo_term = cell_cortex_: extracellular_region_ (0.0)

 nuclear_organization_and_biogenesis_: extracellular_region_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ | NeighGO_term $=$ conjugation_: plasma_membrane_(1.65/0.91) $|\quad| \quad|\quad| \quad|\quad| \quad$ | $|\quad| \quad$ NeighGo_term $=$ oxidoreductase_activity_: extracellular_region_ (0.0) | | | | | | | | | | | NeighGo_term = sporulation_: extracellular_region_ (0.0)
| | | | | | | | | NeighGO term = cytoplasmic_membrane-bound_vesicle_: extracellular_regiōn_(0.0) | | | | | | | | | | | | NeighGO_term = cytoskeleton_: extracellular_region_ (0.0)
 microtubule_organizing_center_: extracellular_region_( $\overline{0} .0$ ) $|\quad| \quad-|\quad| \quad|\quad| \quad|\quad| \quad$ Neigh̄GO_term $=$ signal_transduction_: plasma_membrane_ (4.07/1.39)
| | | | | | | | | | | cellular_respiration_: extracellular_region_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad$ | $\mid \quad$ NeighGo_term $=$ amino_acid_and_derivative_metabolic_process_: extracellūar_region_ (0.0)
| | | | | | | | | | NeighGo_term = endomembrane_system_: extracellular_region_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ $\mid$ NeighGo_term $=$ membrane_organization_and_biogenesis_: extracellular_reḡion_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ $\mid \quad$ Neigh̄GO_term $=$ electron_transport_: extracellular_region_(0.0)
| | | | | | | | | | | protein_catabolic_process_: extracellular_region_(0.0)
| | | | | | | | | | signal transducer activity : extracellular region (0.0)
| | | | | | | | | | | NeighGO_term = cellular_bud_: extracellular region_(0.0)
 site_of_polarized_growth_: extracellular_region_ (0.0)
 cell_budding_: extracellular_region_ (0.0)
| $\mid$ | | | | $\mid$ | | | $\mid$ NeighGo_term $=$
generation_of_precursor_metabolites_and_energy_:
extracellular_region_( $\overline{0} .0$ )
| | | | | $\mid$ | | | | NeighGo_term = phosphoprotein_phosphatase_activity_: extracellular_regīon_(0.0) | | | | | | | | | | | | NeighGO_term = mitochondrial_envelope_: extracellular_region_ (0.0) $|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGO_term $=$ pseudohyphal_growth_: extracellular_region_ (0.0)
$|\quad| \quad|\quad|-|\quad| \quad|\quad| \quad|\quad| \quad$ start > 122944


126324: membrane_(36.23/20.55)
| | | | $\mid$ | | | | | | | | | start $>$ 126324: plasma_membrane_ (12.85)
 | | | | | | | | | | | | | 1 start < = 135930
$\begin{array}{ccccc}\mid & \mid & \mid & \mid & \mid \\ 130640: & \text { cytoplasm_} & \text { (15.18) }\end{array}$
| | | | |
130640
| | | | | | | | | | | | | | | |
start <= 133932: endomembrane_system_(12.47)










```
| | | | | | | | | | | | | | | | |
start <= 327868: vacuole (14.52)
| | | | | | | | |
start > 327868
| | | | | | | | | | | | | | |
| |
Neigh_GO_aspect = F
| | | | | | | | | | |
    neigh_strand = W: nucleus_(4.32/1.5)
    neigh_strand = C: cytoplasm_(4.81/1.69)
| | | | | | | | | | | | | | | | |
Neigh_GO_aspect = P: cytoplasm_(6.64/1.55)
| | | | | | | | | | | | | | start >
333352
| | | | | | | | | | | | | | start <=
335893
| | | | | | | | | | | | | | | |
start <= 334260: mitochondrion_ (18.68)
| | | | | | | | | | | | | | | |
start > 334260: endomembrane_system_ (12.21)
| | | | | | | | | | | | | | | start >
335893: mitochondrion_(42.22/14.88)
| | | | | | | | | | | | | start > 338266
| | | | | | | | | | | | | | start <=
339782: cellular_bud_ (17.69)
| | | | | | | | | | | | | start >
339782
| | | | | | | | | | | | | | |
NeighGO_term = cellular_component_: nucleus_ (3.89/1.6)
| | | | | | | | | | | | | | |
NeighGO_term = hydrolase_activity_: nucleolus_ (0.0)
| | - | | | | \ | | | | | | | | | |
NeighGO_term = biological_process_: nucleolus_ (7.1/1.62)
| | | | | | | | | | | | | | |
NeighGO_term = plasma_membrane_: nucleolus_ (0.0)
| | | | | | | | | | | | | | |
NeighGO_term = transporter_activity_: nucleolus_ (0.0)
| | | | | | | | | | | | | - | | |
NeighGO_term = transport_: nucleolus_ (0.0)
NeighGO_term = endoplasmic_reticulum_: nucleolus_(0.0)
| | | | | | | | | \ | | | | |
NeighGO_term = peroxisome_: nucleolus_ (0.0)
NeighGO_term = protein_binding_: nucleolus_ (0.0)
| | | | | | | | | | | | | | |
NeighGO_term = organelle_organization_and_biogenesis_: nucleolus_
(0.0)
| | | | | | | | | | | | | | |
NeighGO_term = mitochondrion_: nucleolus_ (0.0)
| | | | | | | | | | | | | | | |
NeighGO_term = nucleus_: nucleolus_(4.15/1.23)
```


| | | | | | | | | | | | | | | NeighGo_term = peptidase_activity_: nucleolus_ (0.0)
 NeighGo_term = lipid_metabolic_process_: nucleolus_(0.0)

 NeighGO_term = cellular_homeostasis_: nucleolus_(0.0) | | | | | | | | | | | | | | | NeighGO_term = cell_wall_: nucleolus_(0.0)
| | | | | | | | | | | | | |
NeighGO_term = nucleotidyltransferase_activity_: nucleolus (0.0)

NeighGO_term = meiosis_: nucleolus_ (0.0)
| | | | | | | | | | | | | | |
NeighGO_term = transcription_: nucleolus_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad\rceil \quad|\quad| \quad\rceil|\quad| \quad|\quad|$
NeighGO_term = isomerase_activity_: nucleolus_ (0.0)
| | | | | | | | | | | | | |
NeighGO_term = carbohydrate_metabolic_process_: nucleolus_(0.0)
NeighGo_term = vitamin_metabolic_process_: nucleolus_ (0.0)

NeighGO_term = translation_regulator_activity_: nucleolus_(0.0)
| | | | | | | | | | | | | | |
NeighGO_term = ligase_activity_: nucleolus_ (0.0)
NeighGo_term = membrane_fraction_: nucleolus_(0.0)

NeighGO_term = helicase_activity_: nucleolus_ (0.0)
| | | | | | | | | | | | | | |
NeighGO_term = cell_cortex_: nucleolus_ (0.0)
NeighGo term = cytokinesis : nucleolus (0.0)
| | | | | | | ${ }^{-}$| | | ${ }^{-}$| | | |
NeighGO_term = nuclear_organization_and_biogenesis_: nucleolus_
(0.0)
NeighGO_term = conjugation_: nucleolus_ (0.0)
| | ${ }^{-}$| | | | ${ }^{-}$| | $\left.\right|^{-}$| | | |
NeighGO_term = oxidoreductase_activity_: nucleolus_ (0.0)

NeighGO_term = sporulation_: nucleus_(0.6/0.2)
$\left.|\quad| \quad-\quad|\quad| \quad|\quad|{ }^{-}|\quad| \quad\right\rceil|\quad| \quad|\quad| \quad \mid$
NeighGO_term = cytoplasmic_membrane-bound_vesicle_: nucleolus_(0.0)

NeighGO_term = cytoskeleton_: nucleolus_ (0.0)
NeighGo_term = microtubule_organizing_center_: nucleolus_ (0.0)
$|\quad| \quad-\quad|\quad| \quad|\quad| \begin{array}{lllllll}- & \mid & \mid & \mid & \mid & \mid & \mid\end{array}$
NeighGO_term = signal_transduction_: nucleolus_ (0.0)

| | | | | | | | | | | | | |
NeighGO_term = amino_acid_and_derivative_metabolic_process_: nucleolus_ (0.0)

| | - | | | | ${ }^{-}$| ${ }^{-}$| | ${ }^{-}$| | NeighGO_term = membrane_organization_and_biogenesis_: nucleolus_ (0.0)
| | | | | | | | | | | | | |
NeighGo_term = electron_transport_: nucleolus_(0.0)

| | $\quad$ | | | ${ }^{-}$| | $\quad$ | $\quad$ | $\mid$
NeighGO_term = signal_transducer_activity_: nucleolus_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad|-\mid$
NeighGO_term = cellular_bud_: nucleus_(4.67/0.81)
| | $\quad$ | | | $\mid \quad$ | | | | | |
NeighGO_term = site_of_polarized_growth_: nucleolus_(0.0)

NeighGO_term = cell_budding_: nucleolus_ (0.0)
NeighGO_term = generation_of_precursor_metabolites_and_energy_:
nucleolūs_(0.0)
| | | | | | | | | | | | | | |
NeighGO_term = phosphoprotein_phosphatase_activity_: nucleolus_ (0.0)

$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid$
NeighGO_term = pseudohyphal_growth_: nucleolus_ (0.0)
| | | | | | | | | | | | start > 342517
$\mid$ | | | | | | | | | | start $<=352381$ :
cellular_component_(23.8)
| | $\rceil$ | | ${ }^{-} \mid$| | | | | | | start > 352381 | | | | | | | | | | | | 1 start <=
355451: cytoplasm_(37.52)
| | | | | | | | | | | | | | start >
355451: mitochondrion_(70.78/28.52)
$\begin{array}{lllllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }>360125 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=376657\end{array}$
| | | | | | | | | | | start <= 361244:
cytoskeleton_(21.29)
| | | $\mid$ | | | | | | | | start > 361244
$\mid$ | | | | | | | | | | | start $<=365779$ :
nucleus_(54.78/35.35)
| | | | | | | | | | | | | start > 365779:
cellular_component_ (14.8)
| | | | | | | | | | | start > 376657:
cytoplasm_(26.79)
$\begin{array}{llllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }>378820 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\ \mid & \mid & \mid \text { start }<=411034 \\ \mid & \text { start }<=381322:\end{array}$
endoplasmic_reticulum_(15.76)



| | | | | | | | | | | NeighGO_term = nucleus_: nucleus_(5.66/2.73)
| | | | | | | | | | ribosome_biogenesis_and_assembly_: cytoplasm_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGo_term $=$ RNA_binding_: cytoplasm_(0.0)
| - | - | | - | | | | RNA_metabolic_process_: cytoplasm_(0.0) lyase_activity_: cytoplasm_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGo_term $=$ DNA_metabolic_process_: cytoplasm_(1.67/0.66)
| ${ }^{-}$| $\left.\right|^{-}$| $\left.\right|^{-}$| | $\left.\right|^{-}$| | NeighGo_term $=$ cell_cycle_: cytoplasm_(0.0)
| | | | | | | | | | | | NeighGO_term = molecular_function_: cytoplasm_(5.78/1.61)
| | $\left.{ }^{-}| |^{-}|\quad|\right|^{-}|\quad| \quad \mid \quad$ NeighGo term $=$ vesicle-mediated_transport_: plasma_membrane_(5.86/2.45)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ | $\mid \quad$ NeighGo_term $=$ nucleolus_: cytoplasm_(0.0)
$\mid$ | | | | | | | | | 1 NeighGo_term = ribosome_: cytoplasm_(0.0)
| | $\mid$ | | $\mid$ | | | | | NeighGo_term = translation_: cytoplasm_(0.0)
| | | | | | | | | | | NeighGO_term = structural_molecule_activity_: cytoplasm_(0.0) Golgi_apparatus_: cytoplasm_(0.0)
| |- | $\mid$ | | - | | | NeighGo_term = transferase_activity_: plasma_membrane_ (3.1/1.84)
| | | | | | | | | | $\mid$ NeighGO_term = protein_modification_process_ cytoplasm (5.61/1.3)
। $\quad$ । $\mid$ | $|\quad| \quad \overline{\mid} \mid \quad$ । $\mid$ NeighGo_term $=$ chromosome : cytoplasm (0.0)
 DNA_binding_: cytoplasm_(1.9/1.05)
| | | | | | | | | | 1 NeighGO_term = response_to_stress_: cytoplasm_ (0.0)
| | $\left|-\left|\left.\right|^{-}\right| \quad\right| \quad|-|\quad| \quad$ NeighGo_term $=$ cytoplasm_: cytoplasm_(5.61/2.79)
| | | | | | | | | | NeighGo_term = cytoskeleton_organization_and_biogenesis_: cytoplasm_( $\overline{0} .0)$
 membrane_: cytoplasm_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGo_term $=$ enzyme_regulator_activity_: cytoplasm_ (0.0) transcription_regulator_activity_: cytoplasm_ (0.0)
 extracellular_region_: cytoplasm_(0.0)
| | | | | $\quad$ | | $\quad$ | | cell_wall_organization_and_biogenesis_: cytoplasm_(0.0)
| | | | | | | | | | | vacuole_: cytoplasm_(0.0)
| | | | | | | | | | | | NeighGO term = peptidase_activity_: cytoplasm_(0.0)
| | $\left.\left.\right|^{-}| |^{-}|\quad|\right|^{-}|\quad| \quad \mid \quad$ NeighGo_term $=$ lipid_metabolic_process_: cytoplasm_(0.0)
| | | | | | | | | | | | NeighGO_term = protein kinase activity : cytoplasm (0.0) | | - | $\quad$ | | $\mid$ | | $\mid$ | cellular_homeostasis_: cytoplasm_(0.0)
| | $\mid$ | | $\mid$ | | $\quad$ | | | NeighGo_term $=$ cell_wall_: cytoplasm_(0.0)
 nucleotidyltransferase_activity_: cytoplasm_ (0.0)
| | | | | | | | | | | | NeighGo_term = meiosis_: cytoplasm_ (0.0)
| | | | | | | | | | | NeighGO_term = transcription_: cytoplasm_(3.57/1.97) $\quad$ NeighGO term = isomerase_activity_: cytoplasm_(0.0)
| | | | | | | | | ${ }^{-}$| ${ }^{-}$| carbohydrate_metabolic_process_: cytoplasm_(0.0)
| | | | | | | | | | | NeighGO_term = vitamin_metabolic_process_: cytoplasm_(0.0)
| | | | $\left.{ }^{-}| |^{-}|\quad|\right|^{-} \mid \quad$ NeighGo_term $=$ translation_regulator_activity_: cytoplasm_ (0.0)
| | | | | | | | | | | NeighGO_term = ligase_activity_: cytoplasm_(0.0)
 membrane_fraction_: cytoplasm_(0.0) | | $\mid$ | ${ }^{-}$| | | | | helicase_activity_: cytoplasm_(0.0) | | | | | | | | cell_cortex_: cytoplasm_ (0.0)
| $\mid$ - | | - | | | | cytokinesis_: cytoplasm_(0.0)
| | | $\mid$ | | - | | | | nuclear_organization_and_biogenesis_: cytoplasm_ (0.0)
| | | | | | | | | | NeighGo_term = conjugation_: cytoplasm_(0.0)
| | | - | | - | | | | | NeighGo_term = oxidoreductase_activity_: cytoplasm_ (0.0) | | | | | | | | | | | | NeighGo_term = sporulation_: cytoplasm_(0.0)
| | | | | | | | | | | | NeighGo_term = cytoplasmic_membrane-bound_vesicle_: cytoplasm_ (0.0)
| | | | | | | | | ${ }^{-}$| | ${ }^{-}$NeighGo_term = cytoskeleton_: cytoplasm_(0.0)
| | | $\mid$ | | | | | | | | NeighGo_term = microtubule_organizing_center_: cytoplasm_(0.0)
 signal_transduction_: cytoplasm_(0.0)
| | | | | - | | - | | | NeighGo_term = cellular_respiration_ cytoplasm_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad$ | $\mid$ NeighGo_term $=$ amino_acid_and_derivative_metabolic_process_: cytoplasm_ (0.0)
| | | | | | | | | | | | NeighGO_term = endomembrane_system_: cytoplasm_(0.0) $|\quad| \quad|\quad|-|\quad| \quad$ | $\mid \quad$ | membrane_organization_and_biogenesis_: cytoplasm_(0.0)
 electron_transport_: cytoplasm_ (0.0)
| | $\mid$ | $\left.\right|^{-} \mid$| $\left.\right|^{-} \mid$| $\mid$NeighGo_term $=$ protein_catabolic_process_: nucleus_(2.82/0.79)
| | | | | | | | | | | | NeighGo_term = signal_transducer_activity_: cytoplasm_ (0.0)
 cellular_bud_: cytoplasm_(0.0)
| | $\quad$ । $\mid$ | $\mid$ | | | site_of_polarized_growth_: cytoplasm_(0.0) cell_budding_: cytoplasm_(0.0)
| | | $\mid$ | | | | | generation_of_precursor_metabolites_and_energy_: cytoplāsm_(0.0) | | | | | | | | | | | NeighGO_term = phosphoprotein_phosphatase_activity_: cytoplasm_(0.0)
 mitochondrial_envelope_: cytoplasm_(0.0) | | $\left.\right|^{-}| |^{-}| |^{-}|\quad| \quad \mid \quad$ NeighGo_term $=$ pseudohyphal_growth_: cytoplasm_(0.0)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>545780$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=548752$ : mitochondrion_ |

(54.87/19.46)
| | | | | | | | start > 548752:
cellular_component_ (21.13)
| | | | start > 554842
| | | | | start <= 559151
| start <= 555188: nucleolus_ (20.61)
start > 555188: cytoskeletōn_(69.61/21.97)
start > 559151: membrane_fraction_(44.85/16.66)
| | | start > 559409
| | | | start $<=654232$
| | | | | start $<=623573$
| | | | | | strand = C
| | | | | | start $<=$ 572117: cytoplasm_
(108.26/38.78)

| \| | | | | 1 \| | | start > 572117 |
| :---: | :---: | :---: |
| 1 \| | | \| | | | \| start <= 581609 |
| 1 \| | | $1 \quad 1 \quad 1$ | \| | start <= 576594 |
| 1 \| | | \| | | | \| | | start <= 573974: |
| mitochondrion_ (29.78/13.35) |  |  |
| $\left.\right\|_{(21.17)}$ | 1 \| | \| | | start > 573974: membr |
| \| | | | | 1 \| 1 | \| start > 576594: mitocho |
| (36.97/11.78) |  |  |
| 1 \| | | $1 \quad 1 \quad 1$ | start > 581609 |
| 1 \| | | $1 \quad 1 \quad 1$ | \| | start <= 609769 |
| 1 \| | | $1 \quad 1 \quad 1$ | \| | | start <= 587711 |
| 1 \| | | 1 \| 1 | \| start <= 583726 |







## For Molecular Function











```
| | | | | | | | | | | start > 405583:
molecular_function_(14.56)
```



```
molecular function_(34.38)
| | | | | | | | | | | start > 391306:
oxidoreductase_activity_ (9.63)
| | | | | | | | | | start > 392399
| | | | | | | | | | | start <= 398393:
transcription regulator_activity (8.42)
| | | | | | - | | ` | | start > 398393:
molecular_function_(28.1)
| | | | | | | | | start > 405583: protein_binding_
(14.97)
| | | | start > 413393
| | | | | start <= 449195
| | | | | | start <= 414764:
nucleotidyltransferase_activity_ (18.97)
| | | | | | start > 414764
| | | | | | | start <= 448751
```



```
| | | | | | | | | start <= 421657
| | | | | | | | | | | start <= 419556:
molecular_function_(13.31) | | | | | | | | start > 419556:
protein_binding_ (6.27)
| | | | | | | | | | start > 421657:
protein_kinase_activity_ (13.82)
```



```
| | | | | | | | | | | start <= 432638:
molecular_function_(29.79)
| | | | | | | | | | | start > 432638
| | | | | | | | | | | start <= 436796:
transferase activity (11.18)
| | | - | | | | | | | | start > 436796:
molecular_function_(22.84)
| | | | | | | | | | start > 442546
| | | | | | | | | | | start <= 445916:
protein_binding_ (12.24)
| | | | | | | | | | | start > 445916:
molecular_function_(18.57)
| | | | | | | | strand = W
| | | | | | | | | start <= 432638
| | | | start <= 422931:
molecular_function_ (27.78) | | | | | | | | | start > 422931
| | | | | | | | | | | start <= 425151:
lyase_activity_ (7.48)
| | | | | | | | | | | start > 425151:
molecular_function_(12.36)
```







```
| | | | | | | start > 586061
molecular_function (27.89)
| |lllllll
structural molecule activity (14.95)
| | | | | - | | | start > 608301:
transporter_activity_ (17.61)
| | | | | | | start > 609769: oxidoreductase_activity_
(22.59)
```



```
structural_molecule_activity_ (12.87)
| | | | | | | | | | start > 618227:
ligase_activity_ (15.75)
| | | | | | | | | start > 620747:
transferase_activity (15.92)
| | | | | | | start > 623573
| | | | | | | start <= 633609
| | | | | | | | start <= 624830:
molecular_function_ (21.5) | | | | | | | | start > 624830
| | | | | | | | | start <= 628706
| | | | | | | | | | | start <= 627333:
hydrolase_activity_ (14.71)
| | | | | | | | | | | start > 627333:
molecular function (17.15)
| | | | | | | | | | start > 628706:
hydrolase_activity_ (15.91)
| | | | | | | | | | | | | | | | start <= 640819:
molecular_function_(61.3)
| | | | | | | | | start > 640819
| | | | | | | | | | start <= 644095:
peptidase_activity_ (10.64)
| | | | | | | | | | start > 644095:
molecular_function_(14.41)
| | |
| | | | | | start <= 647601: transporter_activity_
(29.68)
```

```
| | | | | | | start <= 649770: molecular_function
(15.61)
| | | | | | | | start > 649770:
structural_molecule_activity_ (14.02)
| | | | start > 655957
| | | | start <= 669516
| | | | | start <= 667638
| | | | | | | start <= 665013: RNA_binding_ (18.38)
| | | | | | | start > 665013: hydrolase_activity_
(16.14)
| | | | | start > 667638: protein_binding_ (19.71)
| | | | | start > 669516
| | | | | start <= 717574
| | | | | | | start <= 678700
| | | | | | | | start <= 672985:
transferase_activity_ (17.21)
transporter_activity_ (10.93)
| | | | | \ | start > 678700
| | | | | | | | | | | | inart <= \04190 
```




```
DNA_binding_(16.94)
| | | | | | | | | | start > 700876:
molecular_function_(8.89)
| | | | | | | | | start > 704190
| | | | | | | | start <= 705738:
transferase_activity_ (23.14)
| | | | | \ | | | start > 705738:
molecular_function_(18.47)
| | | | | | start > 717574
| | | | | | start <= 725778
| | | | | | | | start <= 719660:
transporter_activity_ (16.27)
```



```
hydrolase_activity_ (22.88)
| | | | | | | | | start > 722809:
molecular_function_(14.59)
| | | | | | | start > 725778
| | | | | | | start <= 732434
| | | | | | | | start <= 729590:
oxidoreductase_activity_ (14.09)
| | | | | | | | | start > 729590:
transporter_activity_ (9.84)
| | | | | | | | start > 732434:
oxidoreductase_activity_ (17.17)
```


## For Biological Process

| start <= 570591
| | start <= 295056





| | | | | | | | | | | | start > 95090: biological_process_(8.74)
| | | | | | | | | | | start > 100888:
lipid_metabolic_process_ (15.34)
| | | | | | | | | | start > 105008:
electron transport (26.22/13.05)
$\begin{array}{lllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }>111153 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\ \text { start }<=126324\end{array}$ | | | | | | | | | | start <= 118821:
biological_process_(14.36)
| | | | | | | | | | | start > 118821:
cell_wall_organization_and_biogenesis_ (18.28)
| | | | | | | | | | start > 126324:
biological_process_(17.49)
$\begin{array}{lllllll}\mid & \mid & \mid & \mid & \mid & \text { start }>137376 \\ \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=147098 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid \\ \mid & \mid & \mid \text { start }<=143289 \\ \mid & \text { start }<=140434:\end{array}$
ribosome_biogenesis_and_assembly_ (20.02)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $-\mid$ | - | $\mid$ | $\mid$ | start $>140434:$ meiosis_ (12.3) |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>143289$ |  |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=146356$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ start $<=145157$ : transport_ |

(18.11)
| | | | | | | | | start > 145157:
response_to_stress_ (20.96)

transport_ (6.2/0.8) | | | | | | | | | distance > 411:
organelle_organization_and_biogenesis_ (3.09/0.34)
$\left.\left.\left.\left.|\quad|\right|^{-}|\quad|\right|^{-}\right|^{-}|\quad| \quad\right|^{-} \mid$neigh_strand $=C$ :
transport_(3.78/0.39)
| | | | | | | | | | neigh_num > 1:
organelle_organization_and_biogenesis_(21.89/7.99)

signal_transduction_ (12.41)
| | | | | | | | | | start > 158185
| | | | | | | | | | distance <= 676: RNA_metabolic_process_(4.26/0.3)
| ${ }^{-}$| $\left.\right|^{-}$| $\left.\right|^{-}$| | | | distance $>676$
$\mid$ | | | | | | | | | | 1 neigh_num $<=1$
$\mid$ | | | | | | | neigh_strand $=$
W: organelle_organization_and_biogenesis_ (2.03/0.73)




```
| | | | | | | | | start <= 317279:
protein_modification_process_ (13.21)
| | | | | | | | | | start > 317279:
nuclear_organization_and_biogenesis_ (9.6)
```



```
vitamin_metabolic_process_ (12.66)
| | | | | | | | start > 323381: transport_ (17.51)
| | | | | | start > 327868
| | | | | | | start <= 338266
```



```
biological_process_ (19.58)
| | | | | | | | | start > 333352:
membrane_organization_and_biogenesis_ (23.29)
| | \ | | | | | | start > 334260
| | | | | | | | start <= 335893: vesicle-
mediated transport_ (9.51)
| | \ | | | | | | start > 335893:
carbohydrate_metabolic_process_ (17.01)
| |llllllll
biological_process_ (18.93)
| | | | | | | | | start > 339782:
ribosome_biogenesis_and_assembly_ (18.91)
| | | | | | | | start > 342517: biological_process_
(30.77)
| | | | start > 352381
| | | | | start <= 359671
| | | | | | strand = C: transport_ (12.5)
| | | | | | strand = W
| start <= 355451: translation_ (46.37)
| start > 355451: cellular_resp\overline{p}iration_
(12.56)
| | | | | start > 359671
| | | | | | start <= 373794
| | | | | | | start <= 361244
| | | | | | start <= 360125: biological_process_
(16.62)
| | | | | | | | start > 360125:
cell_wall_organization_and_biogenesis_ (20.96)
| | | | | | | start > 361244:
nuclear_organization_and_biogenesis (36.57)
| | | | | | s̄tart > 373\overline{794}
| | | | | | start <= 374572: translation_ (20.76)
| | | | | | | start > 374572
| | | | | | | | start <= 376657: biological_process_
(25.82)
| | | | | | | | start > 376657: translation_ (16.94)
| | | start > 378355
| | | | start <= 550504
strand = C
    | start <= 467026
```





| | | | | | | | | | | start > 466201: $\underset{\mid}{\text { RNA_metabolic_process_- (8.37) }} \mid \underset{\mid}{\text { | }}$ | | | | | | | | | | | | | | | start $<=$ 491074: response_to_stress_ (13.23)
| | $\quad$ - | ${ }^{-}$| | | | | | | start > 491074 | | | | | | | | | | | | | start $<=496676$ : biological_process_ (25.19)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>496676$ |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| $497348:$ | DNA_metabolic_process | $(8.97)$ |  |  |  |  |  |  |  |  |  |  |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>$ |  | 497348: biological_process_ (8.2)


| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>503929$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=517877$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid \quad$ start $<=507739:$ |

cellular_homeostasis_(18.65)
$|\quad| \quad|\quad| \quad|\quad| \quad$ start $>$ 507739:
membrane_organization_and_biogenesis_(28.93/9.66)

| | | | | | | | | |
electron_transport_ (23.74)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>526328$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=528690:$ |

RNA_metabolic_process (11.87)
| $\mid$ | $\left.\right|^{-}| |^{-} \mid$| | $\mid$start > 528690:
biological_process_ (11.85)

organelle_organization_and_biogenesis_ (20.19/6.7)
| | | ${ }^{-}$| $\left.\left.\right|^{-}\right|^{-}$| | $\left.\right|^{-}$start > 532055
| | | | | | | | | | start <= 534020 :
cell_cycle_(10.34)
| | | | | | | | | | start > 534020:
biological_process_(17.09)
| | | | | | | | | start > 536049
| | | | | | | | | start <= 542495
| | | | | | | | | 1 start $<=538765$
| | | | | | | | | | 1 start $<=538548$ :
translation_(9.85)
| | | | | | | | | | | | start > 538548:
DNA_metabolic process (6.84)
| $-\mid$ | | $\left.\right|^{-}$| | | | | start > 538765:
translation_(16.0)
| | | | | | | | | start > 542495:
DNA_metabolic_process_(14.57)






## Decision Tree Generated for Chromosome Eleven:

## For Cellular Component

```
start <= 117991
| start <= 22234
    | start <= 6108
    | strand \(=\mathrm{C}\)
        | start <= 2182: cellular_component_ (8.98)
        | start > 2182: ribosome (15.21)
        strand \(=W\) : membrane (15. 29 )
        start > 6108
        start <= 18339
        | start <= 11227: plasma_membrane_ (15.66)
        | start > 11227: cellulā̄_componen̄t_ (30.19)
        start > 18339
        neigh_strand \(=W\) : mitochondrion_ (16.26/4.61)
        neigh_strand \(=\) C: plasma_membrane_(20.04/6.04)
        start > 22234
        start <= 46807
            start <= 39164
            strand \(=C\)
            start <= 30688: cytoplasm_(14.9)
            start > 30688
            | start <= 31694: nucleus_(21.75)
        start > 31694: cytoplasm_(55.31/11.87)
```





```
| | | | | | | | NeighGO_term = cell_budding :
nucleus (0.0)
| | | | | | | | NeighGO_term = response_to_stress_:
nucleus_ (0.0)
(0.0)
| | | | | | | | NeighGO term =
translation_regulator_activity_: nucleus_(0.0)
| | | | | | | | NeighGO_term = helicase_activity_:
nucleus_ (0.0)
| | | | | | | | NeighGO_term =
mitochondrial_envelope_: nucleus_(0.0)
| | | | | | | | N
nucleus_ (0.0)
| | | | | | | | NeighGO_term =
lipid_metabolic_process_: nucleus_ (0.0)
| | | | | | | | NėighGO_term =
membrane organization and biogenesis : \overline{nucleus (0.0)}
| | \ | | | | | | | Neigh̄GO_term =
extracellular_region_: nucleus_ (0.0)
| | | | | | | | NeighGO_term = meiosis_: nucleus_
(0.0)
| | | | | | | | NeighGO term =
phosphoprotein_phosphatase_activity_: nucleus_ (0.0)
| | | | | | | | NeighGO_term = cytoplasmic_membrane-
bound_vesicle_: nucleus_(0.0)
| | | | | | | NeighGO_term = isomerase_activity_:
nucleus_ (0.0)
| | | | | | | | NeighGO_term = motor_activity_:
nucleus (0.0)
| | | | | | | | NeighGO_term =
cellular_respiration_: nucleus_ (0.0)
| | | | | | | | NeighGO_term =
nuclear_organization_and_biogenesis_: nucleus_ (0.0)
| | | | | | | NeighGO_term = cytokinesis_:
nucleus (0.0)
| | | | | | | | NeighGO_term =
microtubule_organizing_center_: nucleus_ (0.0)
| | | | | | | |- NeighGO_term =
generation_of_precursor_metabolites_and_energy_: nucleus_ (0.0)
| | | | | | | start > 94504: cellular_component_
(13.02)
| | | | | | start > 100676
| | | | start <= 114632
| | | | | | | | start <= 112508:
endomembrane_system_(14.07)
| | | | | | | | start > 112508: plasma_membrane_
(17.56)
| | | | | | | start > 114632
| | | | | | | neigh_strand = W: cytoplasm_
(18.73/10.8)
| | | | | | | neigh_strand = C:
cytoplasmic_membrane-bound_vesicle_(15.85/7.7)
| | | | strand = W
```














$||||||\mid$ start $|$ 645994: membrane_ (33.77)

## For Molecular Function

start $<=224099$
$\mid$
$\mid$



```
| | | | | | start > 154996:
structural_molecule_activity_ (23.13)
| | | start > - 161347
| | | | start <= 180784
| | | | | start <= 165935
| | | | | | start <= 162563: molecular_function_ (24.1)
| | | | | | | start <= 164390: isomerase_activity_
(13.13)
| | | | | | start > 164390: molecular_function_
(15.86)
| | | | | start > 165935
| | | | | | start <= 176486
| | | | | | start <= 171134
```



```
oxidoreductase_activity_ (9.32)
| | | | | | | | | start > 166549:
hydrolase_activity_ (17.22)
| | | | | | | | start > 168834:
oxidoreductase_activity_ (15.68)
| | | | | | | start > 171134
| | | | | | | | strand = C:
nucleotidyltransferase_activity_ (12.13)
| | | | | | | | strand = W
| | | | | | | | | start <= 171788:
transporter_activity_ (10.1)
| | | | | | | | | start > 171788:
peptidase_activity_ (16.62)
| | | | | | start > 176486
| | | | | | start <= 178520
| | | | | | | start <= 176786: molecular_function_
(15.3)
| | | | | | | | start > 176786:
structural_molecule_activity_ (18.81)
| | | | | | | start > 178520
| | | | | | | | start <= 179672:
oxidoreductase_activity_ (22.97)
| | | | | | | | start > 179672: hydrolase_activity_
(21.69)
```



```
| | | | | | | | start <= 191441:
peptidase_activity_ (5.7)
| | | | | | | | | start > 191441:
molecular_function_(10.12)
| | | | | | | strand = W
| | | | | | | | start <= 182963:
protein_kinase_activity_ (14.24)
| | | | | | | | start > 182963: molecular_function_
(17.06)
| | | | | | start > 193069
| | | | | | start <= 194865: ligase_activity_
(16.67)
| | | | | | | start > 194865
| | | | | | | | start <= 196027: RNA_binding_
| | | | | | | | | start > 196027: motor_activity_
(12.76)
| | | | | start > 200163 
| | | | | | | | | | | i |tart <= 207891
(24.05)
```



```
(39.64)
```



```
| | | | | | | | | start > 216988:
molecular_function_(23.84)
| | |
| | | | | | | strand = C
| | | | | | | start <= 222544
| | | | | | | | | Neigh_GO_aspect = C:
signal_transducer_activity_ (7.1/2.24)
| | | | | | | | | Neigh GO aspect = F:
protein_kinase_activity_ (5.12/1.01)
| | | | | | | | | Neigh_GO_aspect = P:
protein_kinase_activity_ (9.47/2.89)
| | | | | | | | start > 222544: lyase_activity_
(17.21)
| | | | | | | strand = W: protein_binding_ (14.61)
| start > 224099
| | strand = C
| | | start <= 329810
```






```
| | | | | | | start <= 596464:
structural_molecule_activity_ (15.53)
| | | | | | | start > 596464
| | | | | | | | start <= 598532: molecular_function_
(18.63)
| | | | | | | | start > 598532:
structural_molecule_activity_ (14.53)
| | | | | - | start > 599499
| | | | | | start <= 608588
| | | | | | | | | | | start <= | | | start <= 605060 
peptidase_activity_ (14.04)
| | | | | | | | | start > 603874:
molecular_function_(12.89)
| | | | | | | start > 605060: hydrolase_activity_
(15.88)
| | | | | | start > 608588
| | | | | | | | start <= 626435
| | | | | | | | start <= 615372:
molecular_function_(18.43)
| | | | | | | | | start > 615372
| | | | | | | | | | neigh_strand = W:
structural_molecule_activity_ (17.15/7.25)
| | | | | | | | | | neigh_strand = C:
molecular_function_(9.49/2.63)
| | | | | | | start > 626435
| | | | | | | | start <= 635489:
peptidase_activity_ (15.95)
| | | | | | | | | start > 635489:
molecular_function_(35.53)
| | strand = W-
| | | start <= 611167
| | | | start <= 278767
| | | | start <= 264433
| | | | | | start <= 231871
| | | | | | start <= 226214: DNA_binding_ (21.82)
| | | | | | | start > 226214:
transcription_regulator_activity_ (15.91)
| | | |- | | -start > \overline{2}31871
| | | | | | start <= 256414
| | | | | | | | start <= 235785: molecular_function_
(29.37)
| | | | | | | | start > 235785
| | | | | | | | start <= 242227:
transferase_activity_ (17.78)
| | | | | \ | | | start > 242227
| | | | | | | | | | start <= 248564:
protein_kinase_activity_ (9.1)
| | | | | | | | | | start > 248564:
molecular_function_(10.7)
| | | | | | | start > 256414
| | | | | | | | start <= 260776:
structural_molecule_activity_ (39.37)
| | | | | | | | start > 260776
```







## For Biological Process








| | | | | | | | | | start > 261921: biological_process_ (10.51) | | | | | | | | start > 262993: cellular_respiration_(22.75)

```
(51.74)
```

| $\mid$ | $\mid$ | $\mid$ | start $>273038$ |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid \quad$ strand $=C:$ |

protein_modification_process_(15.03)
| | | | | | | strand = W
$\mid$ | | | | | | | start <= 277925: response_to_stress_
(37.98)
| | | | | | | start > 277925:
carbohydrate_metabolic_process_(32.45/13.11)
| | | | | start $>27 \overline{8} 767$
| | | | start $<=325415$
$\mid$ | | | | start $<=288489$
| | | | | | | start <= 280153
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ distance $<=2100 \quad|\quad| \quad|\quad| \quad$ neigh_strand $=\mathrm{W}$ :
organelle_organization_and_biogenesis_ (10.29/4.13)
$\mid$ | | | | | | | | neigh_strand = C: transport_
(8.92/2.51)
| | | | | | | | distance > 2100:
organelle_organization_and_biogenesis_ (15.01/2.71)
$\begin{array}{lllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }>280153 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\ \mid & & \text { stop }<=282923\end{array}$
| | | | | | | | | start <= 281973:
ribosome_biogenesis_and_assembly_ (16.52)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ start > 281973:
RNA_metabolic_process_(27.7/11.41)
| ${ }^{-}| |^{-}|\quad|^{-}|\quad| \quad \mid \quad$ stop > 282923
$\begin{array}{lllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=284674 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\ \text { start }<=282829:\end{array}$
translation_(12.66)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ start > 282829:
organelle_organization_and_biogenesis_ (15.35)

vesicle-mediated_transport_ (12.13)

ribosome_biogenesis_and_assembly_ (14.21)
| | $\quad$ | | $\quad$ | $\quad$ | start > 288489
| | | | | | | strand = C
| | | | | | | | start $<=314456$
| | | | | | | | | | start <= 299226
| | | | | | | | | | | start <= 293952:
biological_process_(16.28)
| | | | | | | | | | | start > 293952:
RNA_metabolic_process_(12.64)
| | | | | | | | | $\left.\right|_{\text {| }}$
(33.25)
| | | | | | | | | start > 314456:
biological_process_ (35.76)


```
| | | | | | | | | | start > 338397:
organelle_organization_and_biogenesis_ (17.94)
| | | | | | - | - strand = \overline{W}: transcription_ (19.65)
| | | | start > 339086
| | | | | start <= 361471
| | | | | | start <= 349107
| | | | start <= 346408
| | | | | | | | | start <= 340956: cell_cycle_
(21.38)
| | | | | | | | | | | | | | | | | | | | | int > 340956 
biological_process_ (24.42)
| | | | | | | | | | start > 345264: cell_cycle_
(18.06)
| | | | | | | | strand = W:
organelle_organization_and_biogenesis_ (15.24)
| | | | | | | | | | | start > 3\overline{4}6408 
pseudohyphal_growth_(29.94/14.7)
| | | | | | | | start > 348780: biological_process_
(32.56)
| | | | | | start > 349107
| | | | | | | strand = C
| | | | | | | | start <= 356392: cell_budding_
(17.73)
| | | | | | | | start > 356392:
cellular_homeostasis_ (12.83)
| | \ | | ` | strand = W
| | | | | | | | start <= 356392
| | | | | | | | | start <= 353137:
DNA_metabolic_process_ (12.99)
| | | | | | | | | start > 353137:
pseudohyphal_growth_(13.72)
| | | | | | | | | | | | | start > 356392 
(25.1/9.47)
| | | | | | | | | start > 358119: vesicle-
mediated_transport_ (18.33)
| | | | | - start > 361471
| | | | | | strand = C
| | | | | | start <= 406884
| | | | | | | start <= 391922
| | | | | | | | start <= 385401
| | | | | | | | | | start <= 381501:
DNA_metabolic_process_ (14.94)
| | | | | | | | | | start > 381501:
amino_acid_and_derivative_metabolic_process_ (18.86)
| | | | | | | | | start > 385401
| | | | | | | | | start <= 389526:
response_to_stress_ (16.22)
| | | | | | | | | | start > 389526:
DNA_metabolic_process_ (18.52)
```








```
| | | | | | | | | | NeighGO term =
electron_transport_: RNA_metabolic_process_ (0.\overline{0})
| | | | | | ` | | - | NeighGO_term =
nucleotidyltransferase_activity_: RNA_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO_term =
transcription_: RNA_metabolic_process_ (0.0)
| | | | | - | | | | | NeighGO_term = chromosome_:
RNA_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO_term =
cytoskeleton_: RNA_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO_term = cell_cycle_:
RNA_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO_term =
organelle_organization_and_biogenesis_: RNA_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO_term =
protein_kinase_activity_: RNA_metabolic_process_ (0.0)
signal_transducer_activity_: RNA_metabolic_process_ (1.11/0.22)
| | | | | | | | | | NeighGO_term = conjugation_:
RNA_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO_term = DNA_binding_:
RNA_metabolic_process_ (0.0)
| | | |
transferase_activity_: RNA_metabolic_process_(\overline{0}.0)
| | | | | | | | | | NeighGO_term =
carbohydrate_metabolic_process_: RNA_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO_term =
protein_modification_process_: RNA_metabolic_process_(0.0)
| | | | | \ | \ | - | Neigh̆GO_term }\mp@subsup{}{}{-}
cell_budding_: RNA_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO_term =
response_to_stress_: RNA_metabolic_process_ (0.\overline{0)}
| | \ - | | | \ | | - | Ne\overline{ighGO_term = ribosome_:}
RNA_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO_term =
translation_regulator_activity_: RNA_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO_term =
helicase_activity_: RNA_metabolic_process_(0.0)
| | \ | | | - | | | | | NēighGO_term =
mitochondrial_envelope_: RNA_metabolic_process_-(0.0)
| | | | | | | | | | NeighGO_term = cell_cortex_:
RNA_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO term =
lipid_metabolic_process_: transport_ (0.46/0.15)
| | | | | | - | | | - | NeighGO_term =
membrane_organization_and_biogenesis_: RNA_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO_term =
extracellular_region_: RNA_metabolic_process_ (\overline{0}.0)
| | | | | | | | | | NeighGO_term = meiosis_:
RNA_metabolic_process_ (0.0)
| | | | | | | | | NeighGO_term =
phosphoprotein_phosphatase_activity_: RNA_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO_term =
cytoplasmic_membrane-bound_vesicle_: RNA_metabolic_process_ (0.0)
```






## Decision Tree Generated for Chromosome Twelve:

## For Cellular Component







| | | | | | | | | | | | start $<=$ 253862: endoplasmic_reticulum_(15.72)
| | | | | | | | | | | | | start > 253862: cellular_component_(23.36)

(29.58)
| | | | | | | | | start > 273916:
cellular_component_(47.12/13.13)

cytoplasm_(27.08/8.9)
| | | | | | | | | | | | | start > 286821: endomembrane_system_(16.0)

plasma_membrane_(16.72)
| | | | | | | | | | | | | start > 290213:
nucleus (24.38/10.94)

membrane_(33.09/13.46)
$\mid$ | $\mid$ | | | | | | start > 300252: cytoplasm_
(25.21)
| | | | | | | | start > 301990
| | | | | | | start $<=302244$ : nucleus
(41.41/17.24)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>302244$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=316108$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid \quad$ start $<=315732:$ |  |

| | | | | | | | | | | start > 315732:
membrane_(28.71/13.95)
| | | | | | | | | | start > 316108:
mitochondrion_(81.51/10.01)
$\begin{array}{lllllll}\mid & \mid & \mid & \text { start }>322298 \\ \mid & \mid & \mid & \mid & \mid & \text { start }<=346586\end{array}$
| | | | | | start $<=327416$
| | | | | | | | | start <= 323545:
plasma_membrane_(22.02)
| | | | | | | | | start > 323545: vacuole_
(29.43/12.12)

| | | | | | | | | | neigh_strand = C:
nucleus_(4.77/0.69)
$|\mid$ | | | | | | | | | neigh_num > 1:
cytoplasm_(17.47/5.86)








| | | | | | | | | | | ribosome (16.17)
| | $\mid$ | | | | | | | | | start > 626333: nucleus_(35.52/13.94)
 646417
| | | | | | | | | | | | | | |
neigh_strand $=W$ : nucleus_ (16.3/6.6)
| | | | | | | | | | | | | | | |
neigh_strand = C: cytoplasm_(11.75/2.45)
| | | | | | | | | | | | | | | start > 646417: cytoplasm (7.04)
| | | | | | | | | | | | | | start > 660718: mitochondrion_ (8.97)

| | | | | | | | | | | | | | start >

634254
| | | | | | | | | | | | | | 1 start <= 636782: ribosome (15.94)
| | | | | | | | | | | | | | | start > 636782: cytoplasm_(13.47)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ start $>640319$
| | | | | | | | | | | | | start <= 642629: mitochondrion (27.98)
| | | | | | | | | | | | | | start $>$
642629
| | | | | | | | | | | | | | | start <=
646417: nucleus_(13.15)
| | | | | | | | | | | | | | | start >
646417: cytoplasm (25.33)
| | | | | ${ }^{-}$| | | | start > 665004
$\begin{array}{llllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=668893 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=665846:\end{array}$
membrane fraction (10.58)
$|\quad| \quad|\quad|-|\quad| \quad|\quad| \quad$ start $>$ 665846:
Golgi apparatus (12.65)











```
| | | | | | | | | | | | | | NeighGO_term
= cell_cycle_: nucleus_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= generation_of_precursor_metabolites_and_energy_: nucleus_ (0.\overline{0})
| | | | | | | | | | | | | | NeighGO_term
= protein_kinase_activity_: nucleus_ (0.0)
| | | | ` | |- | | - | | | | | NeighGO_term
= cell_wall_: nucleus_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= cellular_homeostasis_: nucleus_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= cytoskeleton_organization_and_biogenesis_: nucleus_ (0.0)
| | | | | | | | | | | | | | | | NeighGO_term
= membrane_organization_and_biogenesis_: nucleus_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= peroxisome_: nucleus_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= lyase_activity_: nucleus_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= membrane_fraction_: nucleus_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= meiosis_: nucleus_ (0.0)
| | | | | - | | | | | | | | | NeighGO_term
= cell_cortex_: nucleus_(0.0)
| | | | | | | | | | | | | | NeighGO_term
= cytoskeleton_: nucleus_(0.0)
| | | | | | | | | | | | | | NeighGO_term
= protein_catabolic_process_: nucleus_ (0.0)
| | | | | - | | - | | | | | | | | NeighGO_term
= microtubule_organizing_center_: nucleus__ (0.0) | | | | | | | | | | | | | | | NeighGO_term
= isomerase_activity_: nucleus_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= vitamin_metabolic_process_: nucleus_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= carbohydrate_metabolic_process_: nucleus_ (0.0) | | | | | | | | | | | | | | | NeighGo_term
= nucleotidyltransferase_activity_: nucleus_(0.0)
= cytokinesis_: nucleus_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= mitochondrial_envelope_: nucleus_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= electron_transport_: nucleus_ (0.0)
| | | | | ` | | - | | | | | | NeighGO_term
= nuclear_organization_and_biogenesis_: nucleus_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= conjugation_: nucleus_(0.0)
| | | | | | | | | | | | | | NeighGO_term
= extracellular_region_: nucleus_(0.0)
| | | | | | | | \ | | | | | NeighGO_term
= cell_budding_: nucleus_ (0.0)
| | | | | | \ | | start > 976053
```




## For Molecular Function







```
| | | | | | | NeighGO_term = cytoplasm_: DNA_binding_
(3.92/1.54)
| | | | | | | NeighGO_term =
protein_modification_process_: transcription_regulator_activity_
(0.0)
| | | | | | | NeighGO term =
ribosome_biogenesis_and_assembly_: Eranscription_regulator_activity_
(0.0)
| | | | | | NeighGO_term = nucleolus_:
transcription_regulator_activity_ (0.0)
| | | | | | | NeighGO_term = nucleus_: DNA_binding_
(15.45/5.99)
| | | | | | NeighGO_term = sporulation_:
transcription_regulator_activity (\overline{0.0)}
| | | | | | | NeighGO_term = ribosome_:
transcription_regulator_activity_(\overline{0.0)}
| | | | | | | Neigh̆GO_term = protein_binding_:
transcription_regulator activity (\overline{0.0)}
| | | | | | | NeighGO_term = response_to_stress_:
transcription_regulator_activity_ (2.82/0.23)
| | | | | | | NeighGO_term = cellular_bud_:
transcription_regulator_activity_(0.0)
| | | | | - | Neigh̄GO term =
site_of_polarized_growth_: transcrip}tion_regulator_activity_ (0.0
| | | | | | | NeighGO_term = pseudohyphal_growth_:
transcription_regulator_activity_ (0.0)
| | | | | | | NeighGO_term = RNA_binding_:
transcription_regulator_activity_(0.0)
| | | | | | - | Neigh̄GO_term = cellular_respiration_:
transcription_regulator_activity_(\overline{0.0)}
| | | | | | | NeighGO_term =
organelle_organization_and_biogenesis_:
transcription_regulator_ac\overline{tivity_ (0.0})
| | | | | | - | Neig\overline{hGO_term = lipid_metabolic_process_:}
transcription_regulator_activity_(\overline{0.0)}
| | | | | | | NeighGO_term =
phosphoprotein_phosphatase_activity_:
transcription_regulator_activity_ (0.0)
| | | | | - | Neig\overline{hGO_term = helicase_activity_:}
transcription_regulator_activity_(\overline{0.0)}
| | | | | | | NeighGO_term =
cell_wall_organization_and_biogenesis_:
transcription_regulator_ac\overline{tivity_ (0.\overline{0})}\mathbf{~}=\mp@code{l}
| | | | | | | NeighGO_term = chromosome_:
transcription_regulator_activity (\overline{0.0)}
| | | | | | | Neigh̄GO_term = DNA_binding_:
DNA_binding_(3.18/1.29)
| | | | | | | NeighGO_term = DNA_metabolic_process_:
transcription_regulator_activity_(\overline{0.0)}
| | | | | - | Neigh̄GO_term =
enzyme_regulator_activity_: transcription_regulator_activity_ (0.0)
| | | | | | | NeighGO_term =
signal_transducer_activity_: transcription_regulator_activity_ (0.0)
```



```
| | | | | | | NeighGO term = membrane_fraction_:
transcription_regulator_activity (\overline{0.0)}
| | | | | | | NeighGO_term = meiosis_:
transcription_regulator_activity_ (0.0)
| | | | | | | NeighGO_term = cell_cortex_:
transcription_regulator_activity_(\overline{0.0)}
| | | | | - | Neig\overline{hGO term = cytoskeleton_:}
transcription_regulator_activity_(0.0)
| | | | | | | NeighGO_term =
protein_catabolic_process_: transcription_regulator_activity_ (0.0)
microtubule_organizing_center_: transcription_regulator_activity
(0.0)
| | | | | | | NeighGO_term = isomerase_activity_:
transcription_regulator_activity_ (0.0)
| | | | | | NeighGO_term =
vitamin_metabolic_process_: transcription_regulator_activity_ (0.0)
| | | | | | | N NeighGOterm =
carbohydrate_metabolic_process_: transcription_regulator_activity_
(0.0)
| | | | | | | NeighGO_term =
nucleotidyltransferase_activity_: transcription_regulator_activity_
(0.0)
| | | | | | NeighGO_term = cytokinesis_:
transcription_regulator_activity_(\overline{0.0)}
| | | | | | | NeighGO_term = mitochondrial_envelope_:
transcription_regulator_activity_ (\overline{0.0)}
| | | | | | NeighGO_term = electron_transport_:
transcription_regulator_activity_(\overline{0}.0)
| | | | | | | NeighGO_term =
nuclear_organization_and_biogenesis_:
transcription_regulator_activity_ (0.0)
| | | | | | NeighGO_term = conjugation_:
transcription_regulator_activity (\overline{0.0)}
| | | | | | NeighGO_term = extracellular_region_:
transcription_regulator_activity_ (0.0)
| | | | | | | NeighGO_term = cell_budding_:
transcription_regulator_activity_(\overline{0.0)}
| | | | | | -start > 174981:
transcription_regulator_activity_ (24.49)
| | | | start > 175226
| | | | | start <= 202591
| | | | | | start <= 193282
| | | | | | | strand = C
| | | | | | | | start <= 183622
| | | | | | | | | start <= 180287:
molecular_function_(30.72)
| | | | | | | | | start > 180287:
hydrolase_activity_ (21.84)
| | | | | | | | start > 183622: molecular_function_
(45.04)
| | | | | | | strand = W
| | | | | | | | start <= 180287
```



```
| | | | | | | | | | start > 234082:
structural_molecule_activity_ (17.81)
| | | | | | | | | start > 237704
| | | | | | | | | | start <= 238814:
molecular_function_(16.3)
| | | | | | | | | | start > 238814:
oxidoreductase activity (15.75)
| | | | | | | start > 242233: molecular_function_
(67.84)
| | | | | start > 250979
| | | | | start <= 263195
| | | | | | start <= 255307: protein_binding_
(33.66/13.13)
| | | | | | | start > 255307
| | | | | | | | start <= 259402:
transferase_activity (15.68) | | | | | | | | | | | | | > 259402: hydrolase_activity_
(9.8)
| | | | | | start > 263195
| | | | | | | | | | | | | | | start <= 275212 
(15.19)
| | | | | | | | start > 267170
| | | | | | | | | start <= 273916:
translation_regulator_activity_ (31.57)
| | | | | | | | | start > 273916:
oxidoreductase_activity_ (14.84)
| | | | | | | start > 275212
| | | | | | | | start<= 278863:
transcription_regulator_activity_ (11.47)
| | | | | | | | start > 278863: molecular_function_
(35.54)
| | | strand = W
| | | | start <= 260980
| | | | | | start <= 255307
| | | | | | start <= 242233:
structural_molecule_activity_ (4.83)
| | | | | | | start > 242233
| | | | | | | start <= 248427: molecular_function_
(14.76)
| | | | | | | | start > 248427
| | | | | | | | | start <= 253862:
oxidoreductase_activity_ (14.18)
| | | | | | | | | start > 253862:
molecular_function_(14.36)
l
```



| | | | | | | | | | | | | start > 433871: RNA_binding_(17.88)
 molecular_function_(30.2/1.19)
| | $\left.\right|^{-}| |^{-}|\quad| \quad|\quad| \quad|\quad| \quad$ start $>444689$
| | | | | | | | | | | | | start <= 447576:
ligase_activity_ (11.55)
| | | | | | | | | | | | | start > 447576 | | | | | | | | | | | | | start $<=$
448315: hydrolase_activity_ (19.83)
| | | | | | | | | | | | | | start >
448315: molecular_function_(9.15)


molecular_function_ (11.96)
| | | ${ }^{-}$| ${ }^{-}$| | | | |
structural_molecule_activity_ (37.54)
| | | | | | | | |
| | | | | | | | | | | distance <= 835:
molecular_function_(35.85/2.45)
| | | | | | | | | | | distance > 835
| | | | | | | | | | | distance $<=2930$
$\mid$ | | | | | | | | | | neigh_strand =
W: hydrolase_activity_ (27.35)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ neigh_strand $=C$
= cellular_component_
| | | | | $\mid$ | | | | | | | | | start $<=$
470406: hydrolase_activity_ (5.95)
470406
| | | |
start $<=483638$
| | | | | | | | | | | | | | | |
start <= 482191: molecular_function_(4.42/1.33)

start > 482191: hydrolase_activity_ (5.17)
| | | | | | | ${ }^{-}$| | ${ }^{-}$| | | | | | |
start > 483638: molecular_function_(3.11/0.36)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ | $\mid$ | $\mid$ | $\mid$ |
= molecular_function_: hydrolase_activity_ (16.09/0.3)


```
| | | | | | | | | | | | | | NeighGO_term
= phosphoprotein_phosphatase_activity_: hydrolase_activity_(0.\overline{0})
| | | | | | | | | | | | | | NeighGO_term
= helicase_activity_: hydrolase_activity_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= cell_wall_organization_and_biogenesis_: hydrolase_activity_(\overline{0}.0)
| | | - | | | \ \ | | - | | | - | Neigh̆GO term
= chromosome_: hydrolase_activity_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= DNA_binding_: hydrolase_activity_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= DNA_metabolic_process_: hydrolase_activity_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= enzyme_regulator_activity_: hydrolase_activity_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= signal_transducer_activity_: hydrolase_activity_ (0.0)
| | \ | | - | | \ | | \ | | | | NeighGO_term
= signal_transduction_: hydrolase_activity_ (0.0)
| | | | | | | | | | | | | NeighGO_term
= transcription_: hydrolase_activity_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= RNA_metabolic_process_: hydrolase_activity_ (0.0)
| | | | | | | | | | | | | | NeighGO term
= amino_acid_and_derivative_metabolic_process_: hydrolase_activity_
(0.0)
| | | | | | | | | | | | | | NeighGO_term
= membrane_: hydrolase_activity_ (0.0)
| | | | | | | | | | | | | | NeighGO term
= vesicle-mediated_transport_: hydrolase_activity_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= Golgi_apparatus_: hydrolase_activity_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= structural_molecule_activity_: hydrolase_activity_ (0.0)
| | | \ | | | | | | | | | - | | - | NeighGO_term
= translation_: hydrolase_activity_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= ligase_activity_: hydrolase_activity_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= peptidase_activity_: hydrolase_activity_ (0.0)
| | | | | \ | | \ | | | | | NeighGO_term
= translation_regulator_activity_: hydrolase_activity_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= endomembrane_system_: hydrolase_activity_ (0.0)
| | | | | | | | | | | | | | | NeighGO_term
= cell_cycle_: hydrolase_activity_ (0.0)
| | | \ | | \ | | | | | | | NeighGO_term
= generation_of_precursor_metabolites_and_energy_:
hydrolase_activity_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= protein_kinase_activity_: hydrolase_activity_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= cell_wall_: hydrolase_activity_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= cellular_homeostasis_: hydrolase_activity_ (0.0)
```

| | | | | | | | | | | | | | NeighGO_term = cytoskeleton_organization_and_biogenesis_: hydrolase_activity_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGO_term = membrane_organization_and_biogenesis_: hydrolase_activity_(0.0)
 = peroxisome_: hydrolase_activity_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ | $\mid$ | $\mid$ | 1 | = lyase_activity_: hydrolase_activity_ (0.0)
| | | | | | | | | | | | | | NeighGO_term $=$ membrane_fraction_: hydrolase_activity_ (0.0)
| | | - | - | | = meiosis_: hydrolase_activity_ (0.0)
| | | | | | | | | | | | | | NeighGO_term = cell_cortex_: hydrolase_activity_ (0.0)
| | | | | | | | | | | | | | NeighGO_term = cytoskeleton_: hydrolase_activity_ (0.0)
| | | | | | | | | | | | | NeighGO_term = protein_catabolic_process_: hydrolase_activity_ (0.0)
| | | | | | | | | | | | | | NeighGO_term = microtubule_organizing_center_: hydrolase_activity_ (0.0)
| | | | | | | | | | | | | | NeighGO_term = isomerase_activity_: hydrolase_activity_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGO_term = vitamin_metabolic_process_: hydrolase_activity_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad N e i g h G O \_t e r m$ = carbohydrate_metabolic_process_: hydrolase_activity_ (0.0)
| | | | $\quad$ | $\mid$ | $\mid$ | | $\mid$ | = nucleotidyltransferase_activity_: hydrolase_activity_ (0.0)
| | | | | | | | | | | | | | NeighGO_term = cytokinesis_: hydrolase_activity_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGO_term = mitochondrial_envelope_: hydrolase_activity_ (0.0)
 = electron_transport_: hydrolase_activity_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGO_term = nuclear_organization_and_biogenesis_: hydrolase_activity_ (0. $\overline{0}$ )
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad N e i g h G O \_t e r m$ = conjugation_: hydrolase_activity_ (0.0)
| | | | $\mid$ | | | | | | | = extracellular_region_: hydrolase_activity_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= cell_budding_: hydrolase_activity_ (0.0)
| | | | | | | | | ${ }^{-}$| | distance > 2930:
molecular_function_(20.67/0.67)







```
| | | | | | | | | | | start <= 903724:
peptidase_activity_ (20.47)
| | | | | | | | | | | start > 903724:
molecular_function_(20.06)
```



```
| | | | | | | | | | | start <= 909965:
oxidoreductase_activity_ (21.2)
| | | | | | | | | | | start > 909965:
molecular_function_(11.68)
| | | | | | | | | | start > 912310:
hydrolase activity (22.65)
| | |- | | | start > 914892
| | | | | | start <= 921596
| | | | | | | | start <= 919019: helicase_activity_
(34.54/17.07)
| | | | | | | | start > 919019:
transcription regulator activity (17.63)
| | | | | | - | starE > 921596
| | | | | | | | start <= 955007
| | | | | | | | | | | | | | | | | | | start <= 931754 <= 925567:
molecular function (17.52)
| | | | | | | | | | start > 925567:
structural_molecule_activity_ (17.75)
| | | | | | | | | | | | | | | | | | | | |tart > 931754 
| | | | | | | | | | | start <= 934251:
molecular function (25.59)
| | | | | | | | | | | start > 934251:
RNA_binding_(14.81)
| | | | | | | | | | start > 937539:
molecular function (49.89)
| | | | | - | | | start > 955007
| | | | | | | | | start <= 958425:
transcription_regulator_activity_ (18.84)
| | | | | | | | | start > 958425
| | | | | | | | | start <= 963782:
molecular function (16.26)
| | | | | | | | | | start > 963782:
peptidase_activity_ (18.15)
| | | | | start > 965894
| | | | | | start <= 1024186
| | | | | | start <= 1017694
| | | | | | | | start <= 973392:
enzyme_regulator_activity_ (14.66)
|
molecular function (37.94)
| | | | | | | | | | start > 1002554
| | | | | | | | | | | start <= 1006375:
phosphoprotein_phosphatase_activity_(16.53)
```



| | | | | | | | start > 425187: molecular_function_(44.62)
(19.81)

| | | | | | | | | | | | | | start $<=$
470406: structural_molecule_activity_ (21.47)
| | | | | | | - | | $\mid$ | | | | | start >
470406
| | | | | | | | | | | | | | | |
distance <= 183
| | | | | | | | | | | | | | | | |
neigh_strand $=W$
| | ${ }^{-}$| | | | | | | | | | | |
| start <= 481874: molecular_function_(4.01/0.12)
| | | | | | | | | | | | | | | | | | |
| start > 481874: structural_molecule_activity_ (10.99/1.3)
| | | | | | | ${ }^{-}|\quad| \quad-|\quad| \quad|\quad| \quad|\quad|$
neigh_strand = C: structural_molecule_activity_ (10.28)
| | | | | | | | | | | | | | |
distance > 183: molecular_function_(51.33/9.6)

structural_molecule_activity_ (16.7)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ | $\mid$ start $>$ 487290:
molecular_function_(21.41)
| | | | | | | | | | | start > 488998:
structural_molecule_activity_ (26.68)
| | | | | | | | | | start > 489350:
molecular_function_(96.68/12.65)
| | | | | | | | | start > 502423

oxidoreductase_activity_(16.56)



| | | | | | | | | | | | | | |
neigh_num <= 1: lyase activity (10.06/3.04)

molecular_function_ (32.14)
| | | | $\left.\right|^{-}$| | | | | | start > 744153:
hydrolase_activity_ (39.42)


766542
| | | | | | | | | | | | | start $<=$ 764808: molecular function (17.7)
| | | | | | | | | | | | | | | start >
764808: DNA_binding_ (26.5/7.85)
| | | | | | | | | | | | | | start >
766542
| | | | | | | | | | | | | | start <= 784354: molecular function (81.03)
| | | | | | | | | | | | | | | start >
784354: nucleotidyltransferase_activity_ (10.96)
 | | | | | | | | | | | | | start <=
786442: protein_binding_ (15.27)
| | | | | | | | | | | | | | start >
786442
| | | | | | | | | | | | | | | start <= 790676
| | | | | | | | | | | | | | |
start <= 787664: molecular_function_ (8.4)
| | | | | | | | | | | | | | | |
start > 787664: signal_transducer_activity_ (16.32)

790676: molecular_function_(18.5)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ start > 797178:
structural_molecule_activity_ (42.62)
| | | | | | | | | | start > 805887
| | | | | | | | | | start < = 822592
| | | | | | | | | | 1 start $<=818609$ | | | | | | | | | | | | | start <= 807385:
molecular_function_(19.81)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ start $>807385$ | | | | | | | | | | | | | | start <= 809997: transferase_activity_ (22.0)
| | | | | - | | | | | | | | start > 809997
815810: molecular_function_(17.09)

```
| | | | | | | | | | | | | | | start >
815810
| | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |
start > 816094: molecular_function_ (12.1)
```



```
RNA_binding_(9.0)
| | | | | | | | | | | | | start > 819312:
transferase_activity_ (13.23)
| | | | | | | | | | | start > 822592
| | | | | | | | | | | | start <= 854061
| | | | | | | | | | | | | | start <= 841330
840320: molecular_function_(73.8)
| | | | | | | | | | | | | | start >
840320: hydrolase_activity (14.19)
| | | | | | | | | | | | | | start > 841330
| | | | | | | | | | | | | | start <=
844281: lyase_activity_ (15.22)
| | | | | | | | | | | | | | start >
844281
| | | | | | | | | | | | | | | start <=
849865
```



```
| | | | | | | | | | | | | | | |
start > 846102: protein_binding_ (18.27)
| | | | | | | | | | | | | | | start >
849865: molecular_function_(37.61)
| | | | | | | | | | | | start > 854061:
structural_molecule_activity_ (13.12)
| | | | | - | star\overline{t > 856441}
| | | | | | start <= 947251
| | | | | | | start <= 907079
| | | | | | | | | | | start <= 893628 
| | | | | | | | | start <= 862713
| | | | | | | | | | | | | start <= 856709:
RNA_binding_(6.97) | | | | | start > 856709:
molecular function_(17.06)
| | | | | | | | | | | | start > 857539
| | | | | | | | | | | | | start <= 859551:
protein_binding_ (13.91)
| | | | | | | | | | | | | start > 859551:
enzyme_regulator_activity_ (19.41)
| | | | | | | | | | | start > 862713
| | | | | | | | | | | | start <= 867353:
transferase_activity_ (14.08)
| | | | | | | | | | | | start > 867353:
molecular_function_(35.81)
```




## For Biological Process

```
| | start <= 552272
| | | start <= 167801
| | | | start <= 121321
| | | | | | | | | | intart <= 46264 
| | | | | | strand = C
| | | | | | | start <= 17956
| | | | | | | | | start <= 16072:
biological_process_ (44.52)
| | | | | | | | | start > 16072:
amino_acid_and_derivative_metabolic_process (4.81)
| | | | | | | | starrt > 17956: biological_process_
(60.37)
| [1 
biological_process_ (8.84)
| | | | | | | | | start > 9836: transport_ (12.97)
| | | | | | | start > 17956: biological_process_
(15.13)
\begin{tabular}{lllllll}
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & start \(>28305\) \\
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & start \(<=36360\) \\
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) \\
\(\mid\) & start \(<=35960\)
\end{tabular}
| | | | | | | | strand = C: biological_process_
(25.7)
l | | | | | | | | | | | | strand=W: transport_(16.99)
(13.88)
| | | | | | | | start > 40665: transport_ (12.78)
| | | | | start > 46264
| | | | | start <= 73408
| | | | | | | start <= 53930
```



```
| | | | | | | | | start <= 47462:
ribosome_biogenesis_and_assembly_ (14.3)
| | | | | | | | | | start > 47462: translation_
(11.79)
| | | | | | | | | start > 49937
| | | | | | | | | | start <= 52589:
protein_modification_process_ (13.38)
```




$\mid$ | | | | $\mid$ start $<=177607$ :
amino_acid_and_derivative_metabolic_process_ (18.59)
$\begin{array}{llllllll}\mid & \mid & \mid & \left.\right|^{-} & \mid & \mid & \mid & \text {start }>177 \overline{6} 07 \\ \mid & \mid & \mid & \mid & \mid & \mid \quad \text { start }<=180287\end{array}$
response_to_stress_ (20.05)
$|\quad| \quad|\quad| \quad \mid \quad$ start $>180287$ :
biological_process_ (12.95)
| | | | $\mid$ start $>187128$

protein_modification_process_ (26.48)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ start $>$ 193282: vesiclemediated_transport_ (33.08/10.23)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |

$||||||||||\mid$ start $|$ = 198785:
biological_process_ (21.01)
$|||||||||||l||$ start $|$ 198785:
cellular_respiration_(17.07)
$\mid$ | $|\quad| \quad|\quad| \quad \mid \quad$ strand $=W$ :
ribosome_biogenesis_and_assembly_ (9.03)
$\begin{array}{lllllllll}\mid & \mid & \mid & \mid & - & \mid & \mid & \text { start }>201316 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & & \end{array}$


 translation_ (13.19)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| r 2591$
 204992
 204226: biological_process_ (31.52)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| :--- | :--- | :--- | :--- |
| $204226:$ | DNA metabolic_process | $(12.97)$ |  |

$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ start $>$ 204992

208762: cell_cycle_(25.59/12.66)
$\mid$ | | | | | | | | | | | | | $\mid$ start $>$ 208762: biological_process_ (22.41)

 meiosis_(43.41/26.84)

| $\|\mid$ | $\mid$ |
| :---: | :---: |
| translation | (23.82) |




biological_process_(34.5)





```
| | | | | | | | | | | | | distance <=
1137: response_to_stress_ (5.2/1.69)
| | | | | | ` | | | | | | | distance >
1137: signal_transduction_(26.31/12.79)
| | | | | | | | | | | start > 371621
| | | | | | | | | | | start <= 374356:
translation_(20.33)
| | | | | | | | | | | | start > 374356:
RNA_metabolic_process_ (37.27)
| |lllllllllll
transport_ (13.0)
| | | | | | | | | | | start > 388221:
biological_process_ (7.31)
| | | | | | | | | | start > 395521
| | | | | | | | | | | start <= 398531:
protein_modification_process_ (16.25)
| | | | | | | | | | | start > 398531:
ribosome_biogenesis_and_assembly_ (10.1)
| | | | | | start > | | | | | | | | | | | | C
| | | | | | | | start <= 410724:
cellular_respiration_ (11.54)
| | | | | | | | | start > 410724:
RNA_metabolic_process_ (17.25)
```



```
| | | | | | | | | | start <= 408446:
lipid_metabolic_process_ (29.5)
| | | | | | | | | | start > 408446:
generation_of_precursor_metabolites_and_energy_ (24.54)
| | | | | | | | | start > 410724:
DNA_metabolic_process_(33.7/16.85)
| | | | start }\mp@subsup{}{}{-
| | | | | start <= 517672
| | | | | | start <= 448315
| | | | | | | start <= 433871
| | | | | | | | strand = C
| | | | | | | | start <= 432169
| | | | | | | | | | start <= 427330:
translation (13.67)
| | | | | | | | | | start > 427330:
cytoskeleton_organization_and_biogenesis_(20.58)
```



```
amino_acid_and_derivative_metabolic_process_ (21.63)
```



```
| | | | | | | | | | start <= 418438
| | | | | | | | | | | start <= 417007:
biological_process_ (22.16)
| | | | | | | | | | | start > 417007:
response_to_stress_ (13.66)
| | \ - | | | | | | | start > 418438
```


| | | | | | | | distance <= 835: biological_process_ (28.46/1.95)
 molecular_function_: response_to_stress_ (6.35) $\left.\left.\left.|\quad|\right|^{-}| |^{-}|\quad|\right|^{-}\right\rceil\left|{ }^{-}\right| \quad|\quad| \quad$ NeighGO_term $=$ biological_process_: response_to_stress_ (5.54)
$\left.\left.\left.|\quad|\right|^{-}| |^{-}|\quad|\right|^{-}\right\rceil\left|{ }^{-}\right| \quad \mid \quad$ NeighGo_term $=$
plasma_membrane_: response_to_stress_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGO_term $=$ transporter_activity_: response_to_stress_ (0.0) $\left.|\quad| \quad|\quad| \quad|\quad| \quad\right|^{-}| |^{-}|\quad| \quad$ NeighGo term $=$ transport : response to stress (0.0)
$\mid$ | | | | | | | | | | $\mid$ NeighGO_term = mitochondrion_: response_to_stress_ (0.0)
 transferase_activity_: response_to_stress_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad|^{-}| |^{-}|\quad| \quad$ NeighGo term $=$ oxidoreductase_activity_: response_to_stress_ (0.0)
 endoplasmic_reticulum_: biological_process_(0.33/0.06)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGO term $=$ transcription_regulator_activity_: response_to_stress_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGo_term $=$ vacuole_: response_to_stress_ (0.0)
 hydrolase_activity_: biological_process_ (8.48/1.56)
 cytoplasm_: response_to_stress_ (0.0)
 protein_modification_process_: response_to_stress_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGo term $=$ ribosome_biogenesis_and_assembly_: response_to_stress (0.0)
$|\quad| \quad|\quad| \quad \mid \quad$ | $|\quad| \quad|\quad| \quad|\quad| \quad$ NēighGO term $=$ nucleolus_: response to_stress (0.0)
 nucleus_: response_to_stress_ (0.0)
$\left.|\quad| \quad|\quad|\right|^{-}|\quad| \quad|\quad| \quad$ | $\mid$ NeighGO_term $=$ sporulation_: response_to_stress_(0.0)
$|\quad| \quad|-|\quad| \quad| \quad|\quad| \quad \mid \quad$ NeighGo_term $=$
ribosome_: response_to_stress_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGO_term $=$ protein_binding_: response_to_stress_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGo_term $=$ response_to_stress_: biological_process_ (5.1/0.94)

cellular_bud_: response_to_stress_ (0.0)
 site_of_polarized_growth_: response_to_stress_ (0.0)
$|\quad \overline{\mid}| \quad|\quad| \quad|\quad| \quad\left|\quad{ }^{-}\right| \quad\left|\quad{ }^{-}\right| \quad$ NeighGO_term $=$ pseudohyphal_growth_: response_to_stress_(0.0) RNA_binding_: response_to_stress_(0.0) | $\left.{ }^{-}\left|{ }^{-}\right| \quad|\quad|\right|^{-}|\quad| \quad|\quad| \quad$ NeighGo term $=$ cellular_respiration_: response_to_stress_ (0.0)

```
| | | | | | | | | | | | | NeighGO_term =
```

organelle_organization_and_biogenesis_: response_to_stress_ (0.0)
$|\quad| \quad|\quad| \quad|\quad|{ }^{-}|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGO_term $=$
lipid_metabolic_process_: response_to_stress_ (0.0)
| | | | | | | | | | | | | NeighGO_term =
phosphoprotein_phosphatase_activity_: response_to_stress_( $\overline{0} .0$ )

helicase_activity_: response_to_stress_ (0.0)
$|\quad| \quad|\quad|-|\quad| \quad|\quad| \quad$ | $|\quad| \quad$ NeighGO_term $=$
cell_wall_organization_and_biogenesis_ response_to_stress - (0.0)

chromosome_: response_to_stress_(0.0) | | | | | | |
DNA_binding_: response_to_stress_ (0.0)
| $\left.-\left.\left|{ }^{-}\right| \quad|\quad| \quad\right|^{-} \mid \quad\right\rceil|\quad| \quad \mid \quad$ NeighGO term $=$
DNA_metabolic_process_: response_to_stress_ (0.0)
$\mid$ | | | | | | | | | $\mid$ NeighGO_term =
enzyme_regulator_activity_: response_to_stress_ (0.0)

signal_transducer_activity_: response_to_stress_ (0.0)
$\left.\left.\left|\left.\right|^{-}\right| \quad\left|\left.\right|^{-}\right|\right|^{-}| |^{-}\right\rangle \quad\left|{ }^{-}\right|$NeighGO_term $=$
signal_transduction_: response_to_stress_ (0.0)
$\left|\left.\right|^{-}\right| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGO_term $=$
transcription_: response_to_stress_ (0.0)

RNA metabolic process : response to stress (0.0)
$\left|{ }^{-}\right| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGo_term $=$
amino_acid_and_derivative_metabolic_process_: response_to_stress_
(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ | $\mid$ NeighGO_term $=$
membrane_: response_to_stress_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ | $\mid$ NeighGO_term $=$
vesicle-mediated_transport_: response_to_stress_ (0.0)

Golgi_apparatus_: response_to_stress_(0.0)
| | | | $\mid$ | $\left.\left.\right|^{-}\right|^{-}|\quad| \quad|\quad| \quad$ NeighGO_term $=$
structural_molecule_activity_: response_to_stress_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGo_term $=$
translation_: response_to_stress_ (0.0)

ligase_activity_: response_to_stress_(0.0)
$\left.\left.\left.\left|\left.\right|^{-}\right|\right|^{-}|\quad|\right|^{-}\right|^{-}|\quad| \quad \mid \quad$ NeighGO_term $=$
peptidase_activity_: response_to_stress_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGo_term $=$
translation_regulator_activity_: response_to_stress_(0.0)
| | | | | | | | | | | | NeighGO_term = endomembrane_system_: response_to_stress_(0.0) $\left.|\quad| \quad|\quad|{ }^{-}|\quad| \quad\right|^{-} \mid \quad$ | $\mid \quad$ NeighGO_term $=$ cell_cycle_: response_to_stress_(0.0)
 generation_of_precursor_metabolites_and_energy_: response_to_stress_ (0.0)
 protein_kinase_activity_: response_to_stress_(0.0) cell_wall_: response_to_stress_(0.0)
$\left.\left.|\quad|\right|^{-}|\quad| \quad\right\rceil-\left|\left.\right|^{-}\right| \quad|\quad| \quad \mid \quad$ NeighGo_term $=$ cellular_homeostasis_: response_to_stress_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ | $\mid$ NeighGO_term $=$ cytoskeleton_organization_and_biogenesis_: response_to_stress_ (0.0) $|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGO_term $=$ membrane_organization_and_biogenesis_: response_to_stress_(0.0) $|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad\left|\quad{ }^{-}\right|-$Neigh $\bar{G} O$ term $=$ peroxisome_: response_to_stress_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGO_term $=$ lyase_activity_: response_to_stress_(0.0)
 membrane_fraction_: response_to_stress_(0.0) $\mid$ meiosis_: response_to_stress_ (0.0)
 cell_cortex_: response_to_stress_(0.0)
 cytoskeleton_: response_to_stress_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGO_term $=$
 microtubule_organizing_center_: response_to_stress_ (0.0) $|\quad| \quad-\left.|\quad|\right|^{-}| |^{-}|\quad| \quad{ }^{-}| |^{-}$NeighGO_term $=$ isomerase_activity_: response_to_stress_ (0.0)
 vitamin_metabolic_process_: response_to_stress_ (0.0)
 carbohydrate_metabolic_process_: response_to_stress_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGO_term $=$ nucleotidyltransferase_activity_: response_to_stress_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGO_term $=$ cytokinesis_: response_to_stress_(0.0)
$\mid$ | | | | | | NeighGO_term $=$ mitochondrial_envelope_: response_to_stress_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ | $\mid$ | $\mid$ | $\mid$ ighGO_term $=$ electron_transport_: response_to_stress_(0.0)
 nuclear_organization_and_biogenesis_: response_to_stress_( $\overline{0} .0$ )
$|\quad| \quad|\quad| \quad \overline{\mid}|\quad| \quad|\quad| \quad \mid \quad$ Neigh̄GO_term $=$ conjugation_: response_to_stress_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGo_term $=$
extracellular_region_: response_to_stress_ (0.0)
| | | | | | | | | | | | | NeighGO_term = cell_budding_: response_to_stress_ (0.0 | $\mid$ | $\mid$ | $\left|{ }^{-}\right|$| $\left.\right|^{-}|\quad|$ distance > 2655: biological_process_(27.24/5.46)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |  |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |  |  |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |  |  |
| $472114:$ | biological_process_(16.6) |  |  |  |  |  |  |  |
| $\mid$ |  |  |  |  |  |  |  |  |

472114

474058

| | | | | | | | | | | | | | | |
start > 481874
| | | | | | | | | | | | | | | | |
start <= 483638: translation (17.71)
| | | | | | | | | | | | | | | | |
start > 483638: biological_process_ (14.17)
| | | | | | | | | | | | | start > 485346:
translation_(10.89)
| | - | | | | | | | | start > 487290:
biological_process (23.0)
| | | | | | | | | | | start > 488998:
translation_(23.76)
| | | | | | | | | | start > 489350:
biological_process_(29.7)
| | | | | | | start > 490407
| | | | | | | start <= 498949
| | | | | | | | | start <= 493885
protein_modification_process_(23.13) $\quad$ start < 493256
biological_process_ (12.72)








```
| | | | | | | | | | | NeighGO term =
site_of_polarized_growth_: carbohydrate_metabolic_process_ (0.0)
| | | | | | \ | | | - | NeighGO_term =
pseudohyphal_growth_: carbohydrate_metabolic_process_ (0.0)
| | | | | | | | | | | NeighGO_term =
RNA_binding_: carbohydrate_metabolic_process_ (0.0)
| | | | | | | | | | | | N
cellular_respiration_: carbohydrate_metabolic_process_ (0.0)
| | | | | | | | | | | NeighGO_term =
organelle_organization_and_biogenesis_: cellular_respiration_
(2.25/1.2\overline{3})
| | | | | | | | | | NeighGO_term =
lipid_metabolic_process_: carbohydrate_metabolic_process_ (0.0)
| | | | | | | | | | | NeighGO_term =
phosphoprotein_phosphatase_activity_:
carbohydrate_metabolic_process_(0.0)
| | | \ | | | | ` | | | NeighGO_term =
helicase_activity_: carbohydrate_metabolic_process_-(0.0)
| | \ | | | | | \ | | - NeighGO}_term =
cell_wall_organization_and_biogenesis_:
carbohydrate_metabolic_process_ (0.0)
| | | | | | | | | | | NeighGO_term =
chromosome_: carbohydrate_metabolic_process_ (0.0)
| | | | | | | | | | | NeighGO_term =
DNA_binding_: carbohydrate_metabolic_process_ (0.0)
| | | | | | | | | | | NeighGO_term =
DNA_metabolic_process_: carbohydrate_metabolic_process_ (0.0)
| | | | | | | | | | | NeíghGO_term =
enzyme_regulator_activity_: carbohydrate_metabolic_process_ (0.0)
| | | | | | | | | | | NeighGO_term =
signal_transducer_activity_: carbohydrate_metabolic_process_ (0.0)
| | | | | | | | | | | NeighGO_term =
signal_transduction_: carbohydrate_metabolic_process_(0.0)
| | | | | - | | | | - | | NNeighGO E}erm 
transcription_: carbohydrate_metabolic_process_ (0.0)
| | | | | | | | | | | NeighGO_term =
RNA_metabolic_process_: carbohydrate_metabolic_process_(0.0)
| | | | | | | | | | | NeighGO_term =
amino_acid_and_derivative_metabolic_process_:
carbo\overline{hydrate_métabolic_process_ (0.0})
| | | | | | | | | | | NeighGO_term =
membrane_: carbohydrate_metabolic_process_ (0.0)
| | | | | | | | | | | NeighGO_term = vesicle-
mediated_transport_: carbohydrate_metabolic_process__(0.0)
| | \ | | ` | | | | | | - NeighGO_term =
Golgi_apparatus_: carbohydrate_metabolic_process_(\overline{0}.0)
| | | | | | | | | | \ Neigh̄GO_term =
structural_molecule_activity_: carbohydrate_metabolíic_process_ (0.0)
| | | | | | | | | | | NeighGO_term =
translation_: organelle_organization_and_biogenesis__(8.89/4.93)
| | | | | | | | | | | NeighGO term =
ligase_activity_: carbohydrate_metabolic_process_(0.0)
| | | | | | | | | | | NeighGO_term =
peptidase_activity_: carbohydrate_metabolic_process_(0.0)
```

| | | | | | | | | NeighGO term = translation_regulator_activity_: carbohydrate_metabōlic_process_ (0.0)
| | | | | | | | | | | NeighGO_term = endomembrane_system_: carbohydrate_metabolic_process _ (0.0) | | | $\mid$ | - | | | | | $\overline{\text { NeighGo_̄erm = }}$ cell_cycle_: carbohydrate_metabolic_process_ (0.0)
| | | | | | | | | | NeighGO_term = generation_of_precursor_metabolites_and_energy_: carbohydrate_metabolic_process_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGo_term $=$ protein_kinase_activity_: carbohydrate_metabolic_prōcess_ (0.0)
| | $\mid$ | ${ }^{-} \mid$| | | ${ }^{-}$| Neigh̄GO_term ${ }^{-}=$ cell_wall_: carbohydrate_metabolic_process_ (0.0)
| | | | | | | | | NeighGO_term = cellular_homeostasis_: carbohydrate_metabolic_process_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGo tērm = cytoskeleton_organization_and_biogenesis_: carbohydrate_metabolic_process_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGO_term =
membrane_organization_and_biogenesis_:
carbohydryate_metabolī̄_prōcess_(0.0)
$|||\quad| \quad| \quad| \quad \mid \quad$ NeighGo_term $=$ peroxisome_: carbohydrate_metabolic_process_ (0.0)
| | | | | | | | | | lyase_activity_: carbohydrate_metabolic_process_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGo_term = membrane_fraction_: carbohydrate_metabolic_process_-(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad{ }^{-}$NeighGO$\quad$ term $=$ meiosis_:
carbohydrate_metabolic_process_ (0.0)
 cell_cortex_: carbohydrate_metabolic_process_(0.0)
| $\mid$ | $|~| ~|~| ~|~| ~|~| ~ N e i g h G O ~ t e r m ~=~$ cytoskeleton_ carbohydrate_metabolic_process (0.0)
 protein_catabolic_process_: carbohydrate_metabolic_process_ (0.0) | | | | | | | | | | NeighGO_term = microtubule_organizing_center_: carbohydrate_metabolic_process_ (0.0)
| | | | | | | | | isomerase_activity_: carbohydrate_metabolic_process_(0.0) | | | | | | | | | NeighGO_term = vitamin_metabolic_process_: carbohydrate_metabolic_process_ (0.0) $|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGO_term $=-$ carbohydrate_metabolic_process_: carbohydrate_metabōlic_process_ (0.0)
| | | | | | | | | | | NeighGO_term = nucleotidyltransferase_activity_: carbohydrate_metabolic_process_ (0.0)
| | | | | | | | cytokinesis_: carbohydrate_metabolic_process_(0.0)
mitochondrial_envelope_: carbohydrate_metabolic_process_(0.0)


```
| | | | | | | | | | NeighGO term =
biological_process_: ribosome_biogenesis_and_assembly_(0.0)
| | | | | | | | | | NeighGO term =
plasma_membrane_: ribosome_biogenesis_and_assembly_ (0.0)
| | | | | | | | | | NeighGO_term =
transporter_activity_: ribosome_biogenesis_and_assembly_ (0.0)
| | | | | \ | | - | | NeíghG\overline{O}term = E}rransport_:
ribosome_biogenesis_and_assembly_ (0.0)
| | | | | | | | | | NeighGO_term =
mitochondrion_: ribosome_biogenesis_and_assembly_ (0.0)
| | | | | | | | | | NeighGO_term =
transferase activity : ribosome biogenesis and assembly (0.0)
| | | | | | | | | | NeíghGO_term =
oxidoreductase_activity_: ribosome_biogenesis_and_assembly_ (0.0)
| | | | | | | | | | NeighGO_term =
endoplasmic_reticulum_: ribosome_biogenesis_and_assembly_ (0.0)
| | | | | | | | | | NeighGO_term =
transcription regulator activity : ribosome biog}enesis and assembly
(0.0)
| | | | | | | | | | NeighGO_term = vacuole_:
ribosome_biogenesis_and_assembly_ (0.0)
| | | | | | | | | | NeighGO_term =
hydrolase activity : ribosome biogenesis and as\overline{sembly (0.0)}
| | | | | | | | | | N
RNA_metabolic_process_(9.53/4.61)
| | | | | | | | | | NeighGO_term =
protein_modification_process_: ribosome_biogenesis_and_assembly_
(0.0)
| | | | | | | | | | NeighGO_term =
ribosome_biogenesis_and_assembly_: RNA_metabolic_process_
(11.58/5.18)
| | | | | | | | | | NeighGO_term = nucleolus_:
ribosome_biogenesis_and_assembly_ (17.88/5.64)
| | \ | | - | | | ` | NeighGO_term = nucleus_:
ribosome_biogenesis_and_assembly_(0.0)
| | | | | | | | | | NeighGO_term = sporulation_:
ribosome_biogenesis_and_assembly_ (0.0)
| | | | | | | | | | NeighGO_term = ribosome_:
RNA_metabolic_process_(6.27/1.6)
| | | | | | | | | | NeighGO_term =
protein_binding_: ribosome_biogenesis_and_assembly_ (0.0)
| | | | | | | | | | NeighGO_term =
response_to_stress_: ribosome_biogenesis_and_assembly_ (0.0)
| | \ - | - | | | | | N Neig\overline{hGO term =}
cellular_bud_: ribosome_biogenesis_and_assembly_ (0.0)
| | | | | | | | | | NeighGO_term =
site_of_polarized_growth_: ribosome_biogenesis_and_assembly_ (0.0)
| | | | | | | | | | NeighGO_term =
pseudohyphal_growth_: ribosome_biogenesis_and_assembly_ (0.0)
| | | | | | | | | | NeighGO_term = RNA_binding_
| | | | | | | | | | | start <= 794794:
RNA_metabolic_process_(11.58/4.6)
| | | | | | | | | | | start > 794794:
ribosome_biogenesis_and_assembly_ (6.93/1.03)
```


| | | | | | | | NeighGO_term = translation_regulator_activity_: ribosome_biogeñesis_and_assembly_ (0.0)
| | | | | | | | | 1 NeighGO_term = endomembrane_system_: ribosome_biogenesis_and_assembly_ (0.0)

ribosome_biogenesis_and_assembly_ (0.0)
| | $\mid$ | | - - | $\mid$ | NeighGO_term =
generation_of_precursor_metabolites_and_energy_:
ribosome_biogenesis_and_assembly_ ( 0.0 )
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGO term $=$
protein_kinase_activity_: ribosome_biogenesis_an̄d_assembly_ (0.0)
$|\quad|{ }^{-}| |^{-}|\quad| \quad|\quad| \quad$ Neigh $\bar{G} O \_t \overline{e r m}=c e \bar{l} \_w a l l_{-}$:
ribosome_biogenesis_and_assembly_ (0.0)
$||\quad| \quad| \quad|\quad| \quad$ | $\mid$ eighGO_term =
cellular_homeostasis_: ribosome_biogenesis_and_assembly_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NēighGO_term =
cytoskeleton_organization_and_biogenesis_:
ribosome_biogenesis_and_assemb̄ ly_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad$ | $\mid \quad$ NeighGO_term =
membrane_organization_and_biogenesis_:
ribosome_biogenesis_and_assembly_ (0.0)
$|\quad| \quad|\quad| \quad-|\quad| \quad|\quad| \quad$ NeighGo_term = peroxisome_:
ribosome_biogenesis_and_assembly_ (0.0)
| | | | | | | | | NeighGO_term =
lyase_activity_: ribosome_biogenesis_and_assembly_ (0.0)
| | | | | | | | | NeighGO_term =
membrane_fraction_: ribosome_biogenesis_and_assēmbly_(0.0)
| | $|~| ~|~| ~|~ † ~| ~ \mid ~-~ N e i \overline{g h G O ~ t e r m ~}=$ meiosis_:
ribosome_biogenesis_and_assembly_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGO_term $=$ cell_cortex_:
ribosome_biogenesis_and_assembly_ (0.0)
| | | | | | | | | | NeighGO term =
cytoskeleton_: ribosome_biogenesis_and_assembly_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGO_term =
protein_catabolic_process_: ribosome_biogenesis_and_assembly_ (0.0)
| | | | | | | | | NeighGO_term =
microtubule_organizing_center_: ribosome_biogenēsis_and_assembly_
(0.0)
| | | | | | | | NeighGO_term =
isomerase_activity_: ribosome_biogenesis_and_assembly_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGo_term $=$
vitamin_metabolic_process_: ribosome_biogenesis_and_assembly_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGO_term $=$
carbohydrate_metabolic_process_: ribosome_biogeñesis_and_assembly_
(0.0)
| | | | | | | | | NeighGO_term =
nucleotidyltransferase_activity_: ribosome_biogenesis_and_assembly_
(0.0)
| | | | | | | | NeighGo_term = cytokinesis_:
ribosome_biogenesis_and_assembly_ (0.0)
$|\quad| \quad|\quad| \quad-|\quad| \quad$ | $\mid \quad$ NeighGO_term $=$
mitochondrial_envelope_: ribosome_biogenesis_and_assembly_ (0.0)

```
| | | | | | | | | | NeighGO term =
electron_transport_: ribosome_biogenesis_and_assembly_(0.0)
| | | | | | | | | | NeighGO_term =
nuclear_organization_and_biogenesis_:
ribosome_biogenesis_and_assembly_ (0.0)
| | | | | | | | | | NeighGO_term = conjugation_:
ribosome_biogenesis_and_assembly_ (0.0)
| | | | | | | | | | NeighGO_term =
extracellular_region_: ribosome_biogenesis_and_assembly_ (0.0)
| | | | | | | | | | NeighGO_term =
cell_budding_: ribosome_biogenesis_and_assembly_ (0.0)
| | | | | | | | | - stārt > 7950̄24
| | | | | | | | | start <= 797430:
translation_(14.63)
| | | | | | | | | start > 797430:
biological_process_(10.05)
| | | | | start > 802396
| | | | | | start <= 831115
| | | | | | | start <= 809997: vesicle-
mediated_transport_ (30.75/11.52)
| | | | | | | start > 809997
| | | | | | | start <= 826412
| | | | | | | | start <= 822592:
biological_process_ (7.91)
| | | | | | | | | start > 822592:
nuclear_organization_and_biogenesis_ (9.8)
| | | | | | | | start > 826412
| | | | | | | | | start <= 828729: transport_
(23.92)
| | | | | | | | start > 828729:
biological_process_ (13.58)
| | | | | | start > 831115
| | | | | start <= 862713
| | | | | | | | start <= 837356:
vitamin_metabolic_process_(42.29/19.44)
| | | | | | | | start > 837356
| | | | | | | | start <= 849123
organelle organization and biogenesis (10.57)
| | | | | | | | | | start > 841330:
response_to_stress_(22.59/10.35)
| | | | | | | | | start > 849123
| | | | | | | | | start <= 857539
start <= 849865:
biological_process_ (9.35)
| | | | | | | | | | | start > 849865:
RNA_metabolic_process_ (9.84)
| | | | | | | | | | start > 857539:
organelle_organization_and_biogenesis_ (16.56)
| | | | | | - | - start > 8\overline{6}2713
| | | | | | | | start <= 873553
| | | | | | | | | start <= 871696:
protein_catabolic_process_(12.68)
```




| | | | | | | | | | | | NeighGO term = chromosome_: meiosis_(0.0) DNA binding : meiosis (0.0)
| | | | | | | | | | | DNA_metabolic_process_: meiosis (0.0)
| $-|\quad|^{-}| |^{-}|\quad|{ }^{-}|\quad| \quad|\quad| \quad$ NeighGo_term $=$ enzyme_regulator_activity_: meiosis_(0.0) $\left|\left.\right|^{-}\right| \quad|\quad| \quad|\quad| \quad$ | $\mid \quad$ NeighGo_term $=$ signal_transducer_activity_: meiosis_(0.0)
 signal_transduction_: meiosis_ (0.0)
| | | | | ${ }^{-}$| | | | | transcription_: meiosis_ (0.0)
| | | | | | | | | | | | NeighGO_term = RNA_metabolic_process_: meiosis_(0.0)
| | | | | | | | | | | amino_acid_and_derivative_metabolic_process_: meiosis_( 0.0 ) $\left|\left.\right|^{-}\right|^{-}|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGo_term $=$ membrane_: meiosis_ (0.0)
$|\quad| \quad|\quad|-|\quad| \quad \mid \quad$ NeighGo_term $=$ vesicle-mediated_transport_: meiosis_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGo_term $=$ Golgi_apparatus_: meiosis_(0.0)
| | | | | | | | | | | NeighGO_term = structural_molecule_activity_: meiosis_ (0.0) translation_: meiosis_ (0.0)
| | | | | | | | | | | ${ }^{-}$NeighGo_term = ligase_activity_: meiosis_ (0.0)
| | | | | | | | | | | | NeighGO_term = peptidase_activity_: meiosis_ (0.0)
| | | | | | | $\mid$ | | | NeighGo_term = translation_regulator_activity_: meiosis_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad$ | $\mid$ NeighGo_term $=$ endomembrane_system_: meiosis_(0.0)
| | | | | | | | | | | | NeighGO_term = cell_cycle_: meiosis_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGo term $=$ generation_of_precursor_metabolites_and_energy_: meiosis_(0.0) | | | | | | | | | | NeighGO_term =
 cell_wall_: meiosis_(0.0)
| $\mid$ | ${ }^{-}$| $\mid$| | | | cellular_homeostasis_: meiosis_(0.0)
 cytoskeleton_organization_and_biogenesis_: meiosis_ (0. $\overline{0}$ )
$|\quad| \quad \mid \quad\rceil \quad|\quad| \quad|\quad| \quad|\quad| \quad$ NeīghGo_term = membrane_organization_and_biogenesis_: meiosis_ (0.0)
| | $|\quad| \quad|\quad| \quad|-|\quad| \quad| \quad$ NeighGo_term $=$
 lyase_activity_: meiosis_ (0.0)
| | | | | | | | | | | | NeighGO_term = membrane fraction : meiosis (0.0) | | $\mid$ | | | | | | | | | NeighGo_term = meiosis_: meiosis_ (0.0)
 cell_cortex_: meiosis (0.0) | $|\quad-|$ | | | | | | | cytoskeleton_: meiosis_ (0.0)
| | | $\quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGo_term $=$ protein_catabolic_process_: meiosis_ (0.0)
 microtubule_organizing_center_: meiosis_ (0.0) | | | | | | | | | | | | NeighGo_term = isomerase_activity_: meiosis_ (0.0) | | | | | | | | | | NeighGO_term = vitamin_metabolic_process_ | | | | | | | | | | | | meiosis (2.4/1.15)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ | $\mid$ neigh num $>1$ : cell_cycle_(8.19/3.64)
 carbohydrate_metabolic_process_: meiosis_(0.0) $\left.|\quad| \quad|\quad| \quad\left|\left.\right|^{-}\right|\right|^{-}|\quad| \quad \mid \quad$ NeighGO term $=$ nucleotidyltransferase_activity_: DNA_metabolic_process_ (2.07/0.7) | | | | | | | | | | | | NeighGO_term = cytokinesis_: meiosis_ (0.0)
 mitochondrial_envelope_: meiosis_(0.0)
| | | | | | - | $\quad$ | | | NeighGo_term $=$ electron_transport_: meiosis_(0.0)
| | | | | | | | | | | | NeighGO_term = nuclear_organization_and_biogenesis_: meiosis_(0.0)
| | $|\quad| \quad|\quad \overline{\mid}| \quad \mid \quad$ - $|\quad|-$ NeighGo_term $=$ conjugation : meiosis (0.0)
| | | | | | | | | | | ${ }^{-}$NeighGo_term =

cell_budding_: meiosis_ (0.0)
| $\mid$ | | | $-\quad|\quad| \quad|\quad|$ start > 786442:
conjugation_(8.4)







```
| | | | | | | | start<= 1055068:
biological_process_ (41.57)
| | | | | | | | start > 1055068
| | | | | | | | | start <= 1057332:
protein_modification_process_ (11.9)
| | | | | \ | \ | start > 1057332:
biological_process_ (40.57)
| | | | | | | start > 1066570: DNA_metabolic_process_
(18.99)
```


## Decision Tree Generated for Chromosome Thirteen:

## For Cellular Component

```
| start <= 413981
| | | strand = C
| | | | start <= 21700
| | | | | start <= 14754
| | | | | | start <= 13175
| | | | | | | start <= 7244: mitochondrion_ (5.62)
| | | | | | start > 7244: endoplasmic_reticulum_
(24.39)
| | | | | | start > 13175
| | | | | | | distance <= 1786: mitochondrion_
| | | | | | | | distance > 1786: mitochondrial_envelope_
(31.74/9.87)
| | | | | start > 14754: endoplasmic_reticulum_
(75.53/13.42)
| | | | start > 21700
| | | | start <= 23684: microtubule_organizing_center_
(13.88)
| | | | | start > 23684: plasma_membrane_(16.51)
| | | strand = W
| | | | start <= 13175
| | | | | start <= 7244: membrane_fraction_ (11.48)
| | | | start > 7244: cytoplasm- (7.55)
| | start > 25801
| | | start <= 318679
| | | | start <= 291133
| | | | | start <= 164176
| | | | | | strand = C
| | | | | | | | start <= 56773
| | | | | | | | | | start <= 51640
| | | | | | | | | | start <= 41794
mitochondrion_(10.18)
| | | | | | | | | | | | start > 34243:
Golgi_apparatus_ (14.87)
```







```
| | | | | | | | | | | | | start <=
217362
| | | | | | | | | | | | | | | |
neigh_strand = W
distance <= 1876: membrane (16.45/7.85)
| | | | | | | - | | | | | | | | |
distance > 1876: endomembrane_system_ (13.33/5.71)
| | | | | | | | | | | | | | | |
neigh_strand = C
neigh_num <= 1: endoplasmic_reticulum_ (10.35/4.33)
| | | | | | | - | | | | | | | | | |
neigh_num > 1: membrane_(9.57/2.87)
| | | | | | | | | | | | | | start >
217362: cytoplasm_ (12.54)
| 
mitochondrion_(22.72)
| | | | | | | | | | | | start > 225889
| | | | | | | | | | | | | start <= 228937:
nucleus_(12.81)
| | | | | | | | | | | | | start > 228937:
mitochondrion_(35.3/13.21)
| | | | | | | | | | | strand = W
| | | | | | | | | | | | start <= 225889
| | | | | | | | | | | start <= 221406:
nucleus (11.42)
| | | | | | | | | | | | | start > 221406:
ribosome_ (22.45)
| | \ | | | | | | | | | start > 225889:
nucleus_(20.38/8.21)
| | | | | | | | | start > 230813
| | | | | | | | | | start <= 235952
| | | | | | | | | | | start <= 231149:
cellular_component_ (19.13)
| | | | | | | | | | | start > 231149
| | | | | | | | | | start <= 233457:
endoplasmic_reticulum_ (30.97/11.75)
| | | | | | | | | | | | start > 233457:
vacuole_(40.04/18.02)
| | | | | | | | | | start > 235952
| | | | | | | | | | | start <= 241536:
cytoplasm_(54.1/10.38)
| | | | | | | | | | | start > 241536
| | | | | | | | | | | | start<= 243029:
cellular_component_ (28.92/9.9)
| | | | | | | | | | | | start > 243029:
nucleus (28.18/13.78)
|
endoplasmic_reticulum_(19.77)
```





```
| | | | | | | | NeighGO term = response to stress :
mitochondrial_envelope (0.0)
| | | | | | | | NeighGO_term =
transferase_activity_: mitochondrial_envelope_ (0.0)
| | | | | | | | NeighGO_term =
lipid_metabolic_process_: mitochondrial_envelope (0.0)
| |- | | | | - | | NeighGO_term = cỳtoplasm_:
mitochondrion_(11.18/5.67)
| | | | | | | | NeighGO_term = hydrolase_activity_:
mitochondrial_envelope_ (0.0)
| | | | | | | NeighGO_term =
DNA_metabolic_process_: mitochondrial_e\overline{nvelope (0.0)}
| | | | | | | | Neigh\overline{GO_term = transcription_:}
mitochondrial_envelope_(0.0)
| | | | | | | | NeighGO_term = mitochondrion_:
mitochondrial_envelope_(0.0)
| | | | | | | | NeighGO_term = electron_transport_:
mitochondrial envelope (0.0)
| | | | | | | | NeighGO_term = biological_process_:
mitochondrial_envelope_(12.9/6.81)
| | | | | | | | NeighGO_term = Golgi_apparatus_:
mitochondrial_envelope_(0.0)
| | | | | | | | NeighGO_term = DNA_binding_:
mitochondrial_envelope_(0.0)
| | | | | | | | NeighGO_term =
cellular_respiration_: mitochondrial_envelope_ (0.0)
| | \ | | \ | | Neigh̆GO_term = ribosome_:
cellular_component_(8.68/2.77)
| | | | | | | | NeighGO_term = transport_:
mitochondrial_envelope_ (0.0)
| | | | | | | | NeighGO_term = chromosome_:
mitochondrial_envelope_ (0.0)
| | | | | | | | NeighGO_term =
transcription_regulator_activity_: mitoc
| | | | | | | | NeighGO_term =
amino_acid_and_derivative_metabolic_process_:
mitochondrial_envelope_ (0.0)
| | | | | | | | NeighGO_term = cell_cycle_:
mitochondrial_envelope_(5.09/2.84)
| | | | | | | | NeighGO_term =
enzyme_regulator_activity_: mitochondrial_envelope_(0.0)
| | | | | | | | NeighGO_term = protein_binding_:
mitochondrial_envelope_ (0.0)
| | | | | | | | NeighGO_term = peptidase_activity_:
mitochondrial envelope (0.0)
| | | | | | | | NeighGO_term =
protein_catabolic_process_: mitochondrial_envelope_(0.0)
| | | | | | | | NeighGO_term = lyase_activity_:
mitochondrial_envelope_ (0.0)
| | | | | | | | NeighGO term =
transporter_activity_: membrane_(4.2/2.23)
| | | | | | | NeighGO_term =
generation_of_precursor_metabolites_and_energy_:
mitochondrial_envelope__(0.0)
```










```
| | | | | | | | | NeighGO term =
molecular_function_: cytoplasm_(0.0)
| | | | | | | | | NeighGO_term =
response_to_stress_: cytoplasm_ (0.0)
| | | | | | | | | NeighGO_term =
transferase_activity_: cytoplasm_ (0.0)
| | | - | \ | | \ NeighGO_term =
lipid_metabolic_process_: cytoplasm_ (0.0)
| | | | | | | | | NeighGO_term = cytoplasm_:
cytoplasm_(0.0)
| | | | | | | | | NeighGO_term =
hydrolase_activity_: cytoplasm_ (0.0)
| | | | | | | | | NeighGO_term =
DNA_metabolic_process_: cytoplasm_ (0.0)
| | | | | | | | NeighGO_term = transcription_:
cytoplasm (0.0)
| | | | | | | | NeighGO_term = mitochondrion_:
cytoplasm_(8.05/2.69)
| | | | | | | | NeighGO_term =
electron_transport_: cytoplasm_ (0.0)
| | | | | | | | | NeighGO_term =
biological_process_: cytoplasm_(0.0)
| | | | | | | | | NeighGO_term = Golgi_apparatus_:
cytoplasm_(0.0)
| | | | | | | | | NeighGO_term = DNA_binding_:
cytoplasm_(0.0)
| | | | | | | | | NeighGO_term =
cellular_respiration_: cytoplasm_(0.0)
| | | | \ | | \ NeighGO_term = ribosome_:
cytoplasm_ (0.0)
| | | | | | | | | NeighGO_term = transport_:
cytoplasm_ (0.0)
| | | | | | | | | NeighGO_term = chromosome_:
cytoplasm_ (0.0)
| | | | | | | | | NeighGO_term =
transcription_regulator_activity_: cytoplasm_ (0.0)
| | | | | | | | | NeighGO_term =
amino_acid_and_derivative_metabolic_process_: cytoplasm_ (0.0)
| | | | | | | | | NeighGO_term = cell_cycle_:
cytoplasm_(0.0)
| | | | | | | | | NeighGO_term =
enzyme_regulator_activity_: cytoplasm_(0.0)
| | | | | | | | | NeighGO_term = protein_binding_:
cytoplasm_(0.0)
| | | | | | | | NeighGO_term =
peptidase_activity_: mitochondrion_(10.2/4.35)
| | | | | | | | | NeighGO_term =
protein_catabolic_process_: cytoplasm_(0.0)
| | | | | | | | NeighGO_term = lyase_activity_:
cytoplasm_(0.0)
| | | | | | | | NeighGO_term =
transporter_activity_: cytoplasm_ (0.0)
| | | | | | | | | NeighGO_term =
generation_of_precursor_metabolites_and_energy_: cytoplasm_ (0.0)
```











```
| | | | | | | | | NeighGO_term = ribosome_:
ribosome (0.0)
| | \ | | | | | | NeighGO_term = transport_:
|ibosome_ (0.0) | | | | | | NeighGO_term = chromosome_:
ribosome_(0.0) | | | | | | NeighGO term=
transcription_regulator_activity_: nucleolus_ (1.44/0.59)
| | | | | | | | | NeighGO_term =
amino_acid_and_derivative_metabolic_process_: ribosome_ (0.0)
| |
ribosome_ (0.0)
| | \ | | | | | | NeighGO_term =
enzyme_regulator_activity_: ribosome_ (0.0)
| | | | | | | | | NeighGO_term = protein_binding_:
ribosome_(0.0)
| \ | | | | | | NeighGO_term =
peptidase_activity_: ribosome_ (0.0)
| | | | | | | | | NeighGO_term =
protein_catabolic_process_: ribosome_ (0.0)
| | | | | | | | | NeighGO_term = lyase_activity_:
ribosome (0.0)
| | | | | | | | NeighGO_term =
transporter_activity_: ribosome_ (0.0)
| | | | | | | | | NeighGO_term =
generation_of_precursor_metabolites_and_energy_: ribosome_(0.0)
| | | | | | | | | NeighGO_term =
cytoplasmic_membrane-bound_vesicle_: ribosome_(0.0)
| | | - | | | - | | - NeighGO_tērm = vesicle-
mediated_transport_: ribosome_ (0.0)
| | | | | | | | | NeighGO_term =
microtubule_organizing_center_: ribosome_(\overline{0}.0)
| | | | | | | | | NeighGO_term =
structural molecule_activity_: ribosome (0.0)
| | | | | - | | ` | Neig}hGO_term = translation_:
ribosome (0.0)
| | | | | | | | | NeighGO_term = membrane_:
ribosome_(0.0)
| | | | | | | | | NeighGO_term = cell_cortex_:
ribosome_ (0.0)
| | \ | | | | | | NeighGO_term =
membrane_fraction_: ribosome_ (0.0)
| | | | | | | | | NeighGO_term = plasma_membrane_:
ribosome_ (0.0)
| | \ | | | | | | NeighGO_term = RNA_binding_:
ribosome_(4.66/1.7)
| | \ | | | | | | NeighGO_term =
RNA_metabolic_process_: ribosome_(3.51/0.55)
| | | | | | | | NeighGO_term = conjugation_:
ribosome_ (0.0)
| | | | | | | | | NeighGO_term = peroxisome_:
ribosome_ (0.0)
| | | | | | | | | NeighGO_term =
vitamin_metabolic_process_: ribosome_ (0.0)
```

| | | | | | | | NeighGO_term = endomembrane_system_: ribosome_(0.0)
| | | $\mid$ | | | | NeighGO_term = cytoskeleton_organization_and_biogenesis_: ribosome_(0.0)
 ribosome_ (0.0)
| | $\mid$ | | | | NeighGO term = phosphoprotein_phosphatase_activity_: ribosome_ (0.0)
| | | | | | | | | NeighGO_term = cellular_homeostasis_: ribosome_ (0.0)
 signal_transduction_: ribosome_ (0.0)
| | | | | | | | | NeighGO_term = protein_kinase_activity_: ribosome_ (0.0)
| | | | | | | | NeighGO_term = extracellular_region_: ribosome_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad$ | NeighGo_term = nucleolus_: ribosome (7.69/1.05)
| | $\mid$ | | | | | NeighGO_term =
ribosome_biogenesis_and_assembly_: ribosome_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad$ $\mid$ NeighGO_term = sporulation_:
ribosome_ (0.0)
$|\quad| \quad|\quad| \quad \mid \quad$ NeighGO_term = organelle_organization_and_biogenesis_: ribosome_ (0.0)

ribosome_ (0.0)
| | $\mid$ | | | | | NeighGO_term =
site_of_polarized_growth_: ribosome_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGO_term = cytokinesis_: ribosome_ (0.0)
| | $\mid$ | | | | $\mid$ NeighGO_term = mitochondrial_envelope_: ribosome_ (0.0)
| | | | | | | | NeighGO_term =
nuclear_organization_and_biogenesis_: ribosōme_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGo_term = meiosis_:
ribosome_(0.0)
| $\quad$ | $\mid$ | | | | NeighGO_term = membrane_organization_and_biogenesis_: ribosome_ (0.0)
| | $\mid$ | | | | cell_wall_organization_and_biogenesis_: ribosome_ (0.0)
| | | | | | | | NeighGO_term =
helicase_activity_: ribosome_ (0.0)
| | $\mid$ | | | | $\mid$ | NeighGO_term = isomerase_activity_: ribosome_ (0.0)
| | $\left.\left.\right|^{-}| |^{-}|\quad|\right|^{-} \mid$NeighGO_term =
carbohydrate_metabolic_process_: ribosome_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighḠ_term = cell_budding_:
ribosome_ (0.0)
| | $\rceil$ | | | | | | NeighGO_term = ligase_activity_:
ribosome_ (0.0)
| | $\mid$ | | | | | NeighGO_term = motor_activity_:
ribosome_(0.0)
$|\quad| \quad|\quad| \quad \mid \quad$ NeighGO_term =
anatomical_structure_morphogenesis_: ribosome_ (0.0)




## For Molecular Function



| | | | | | | | | | | | start $<=$ 95369: molecular_function_(11.25)
| | | | | | | | | | |
oxidoreductase_activity_ (12.24) | | | | | | | | |
$\mid$ | | | | | | | | | | start $<=99400$ : structural molecule_activity_(6.12)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ | $\mid$ start $>99400$ :
molecular_function_(15.97)
| | | | | | | | | | start > 101862
| | | | | | | | | | start <= 108806:
transporter_activity_ (29.38)
DNA_binding_(36.54/16.1)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>115347$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=121324$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=118898:$ |  |

oxidoreductase_activity_(22.44)
| | | | | | | | | start > 118898:
isomerase_activity_ (15.02)

| | | | | | | | | | | | start $<=$ 142210:
molecular_function_(62.36)
| | $\left.\right|^{-}| |^{-}|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ start $>142210$

146482: protein_binding_(12.8)
| | | | | | | | | | | | | | start > 146482: molecular function_(15.71) | | | | | | | ${ }^{-}$| | | helicase_activity (15.34)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>151871$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=167308$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| $\mid$ | $\mid$ | start $<=163620$ |  |  |  |  |  |  |  |  |  |
| start $<=160180:$ |  |  |  |  |  |  |  |  |  |  |  | hydrolase_activity_ (18.31)

| | | | | | | | | | | | start > 160180: RNA_binding_(14.63
| | | | | | | | | | | start > 163620:
oxidoreductase_activity_ (37.14)
| | | | | | | | | | 1 start > 167308
| | | | | | | | | | | start <= 173139:
molecular_function_(28.88)
| | | | | | | | | | | | start > 173139
| | | | | | | | | | | | | start $<=$ 178426:
RNA binding (19.57)
| | | | | | | | | | | | start > 178426:
molecular_function_(9.34)
| | | | | | | strand = W
| | | | | | | | start <= 159383


| | | | | | | | | start <= 241536: phosphoprotein_phosphatase_activity_ (21.07)
| | | | $\left.{ }^{-}|\quad|\right|^{-}| |^{-} \mid$start $>241536$
transcription_regulator_activity_ (13.53)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>244149$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=251516$ |

molecular function_(18.74)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ start > 247677:
structural_molecule_activity_ (11.81)
| | | | | | | | | | start > 251516
| | | | | | | | | | | start $<=$ 252990:
transferase_activity_ (24.85)

molecular_function_(11.62)


| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>196170$ |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=214189$ |

| | | | | | | | | | | | start <= 209525:
molecular_function_(21.69)
| | | | | | | | | | | | start > 209525:
structural_molecule_activity_ (7.21)
$|\quad| \quad|-|\quad| \quad| \quad|\quad| \quad$ start $>$ 214189:
molecular_function_ (33.27)
| | | | | | | | | start > 217362
| | | | | | | | | | start <= 220138:
protein_binding_ (14.04)
| | | | | | | | | | | start > 220138
| | | | | | | | | | | start $<=$ 221406:
DNA_binding_(14.48)
| | | | | | | | | | | | start > 221406:
structural_molecule_activity_ (9.95)
| | | | | - | $|\quad| \quad$ start > 225889
$\mid$ | | | | | | | start $<=243225$
| | | | | | | | | 1 start $<=236588$
| | | | | | | | | | | start <= 228937:
transferase_activity_ (13.9)
| | | | | | | | | | | | start > 228937
| | | | | | | | | | $\mid$ start $<=231149$ :
molecular_function_(15.21)
| | | $\left.\right|^{-}$| $\left.\right|^{-}|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ start $>231149$
$\mid$ | | | | | | | | | | | | 1 start <=
233457: transferase_activity_ (14.84)
| | | | | | | | | | | | | | start >
233457: molecular_function_(12.8)
| | | | | | | | | | start > 236588:
transferase_activity_ (23.32)



```
| | | | | | | | start <= 315376:
oxidoreductase_activity_ (26.44)
| | | | | | | | start > 315376
| | | | | | | | | | start <= 318679:
molecular_function_(13.53)
| | | | | | | | | start > 318679:
structural_molecule_activity (13.73)
| | | | | - | start > 321874
| | | | | | start <= 337787
| | | | | | | | | | | | | | | | start <= 337312 
molecular function (38.26)
```




```
protein_binding_ (17.92)
| | | | | | | | | | | | start > 327481:
molecular_function_ (18.35)
| | | | | | | | | | | start > 330792:
protein_binding_ (23.47)
| | | | | | | | | | start > 335297:
molecular_function_(12.55)
| | | | | | | | start > 337312:
transcription_regulator_activity_ (12.3)
| | | | | | | start > 337787
| | | | | | | start <= 343519: peptidase_activity_
(14.69)
\begin{tabular}{lllllllll}
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & start \(>343519\) \\
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) \\
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid \quad\) start \(<=353870\) \\
& & & &
\end{tabular}
molecular_function_(32.26)
| | | | | | | | | | start > 352279: DNA_binding_
(27.21)
| | | | | | | | | start > 353870
| | | | | | | | | | start <= 355383:
protein_binding_ (11.72)
| | | | | | | | | | start > 355383:
molecular function (41.37)
| | | | start > 382897
| | | | | start <= 392754
| | | | | | start <= 391098
| | | | | | strand = C
enzyme_regulator_activity_ (19.5)
| | | | | | | | start > 387020:
transporter_activity_ (19.66)
| | | | | | | strand = W
| | | | | | | | start <= 383302:
transporter_activity (40.92/18.79)
| | | | | | | | start > 383302
| | | | | | | | | start <= 388821:
oxidoreductase_activity_ (14.38)
```




| | | | | | | | start > 571015: molecular_function_(65.74)
| | | | | | | start > 577717: hydrolase_activity_
(19.08)

molecular_function_(27.08)

| | | | | | | | | | | start < = 592627:
hydrolase_activity_ (17.82)
| | | | | | | | | | | start > 592627:
transporter activity (13.44)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ start $>594885$
| | | | | | | | | | | start <= 599157:
DNA_binding_(13.56)
| | | | | | | | | | start > 599157:
oxidoreductase_activity_ (34.33)

molecular_function_ (16.25) | | | | | | |
enzyme_regulator_activity_ (16.34)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>616565$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| start $<=623212:$ |  |  |  |  |  |  |  |  |  |  |  |

hydrolase activity (11.4)
| | | ${ }^{-}$| | | | | | $\left.\right|^{-}$start > 623212:
molecular_function_(14.99)
| | | | | | | | | start > 624079:
transcription_regulator_activity_ (14.86)
| | | | | | | start > 626348
| | | | | | | start $<=667043$
| | | | | | | | start $<=655075$
| | | | | | | | | start <= 629024:
protein_binding_ (10.45)
| | | | | | | | | start > 629024
| | | | | | | | | start $<=637499$
| | | | | | | | | | | start $<=$ 635983:
molecular_function_(17.48)
| | | | | | | | | | | | start > 635983:
structural_molecule_activity_ (16.71)
| | | | | | | | | | | start > 637499













## For Biological Process












$\mid$ | | | | | | | | $\mid$ start > 335297: biological_process_ (15.57)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>337312$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=337787:$ |

RNA_metabolic_process_(22.66/10.08)
| | | | | | | | | | | start > 337787:
protein_modification_process_ (21.42)
| | | | $|\quad| \quad$ | start > 343519
$\begin{array}{llllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=352602 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\ \mid & \mid \quad \text { start }<=352279 \\ \text { start }<=350380:\end{array}$
biological_process_(24.59)
| | | | | | | | | | | start > 350380:
translation_(18.81)
$\begin{array}{llllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }>352279 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \text { distance }<=2848:\end{array}$
amino_acid_and_derivative_metabolic_process_ (18.55/7.67)
| $\left.\left.\left.\left.\right|^{-}\right|^{-}\right|^{-}|\quad|\right|^{-}|\quad| \quad \mid \quad$ distance $>$ 2848:
biological_process_(7.1/2.08)
| | | | | | | | $\left.\right|^{-}$| 1 |art > 352602
$|||||||\mid$ start $<=355383$ :
DNA_metabolic_process_ (24.42)
$\begin{array}{llllllllll}- & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }>355383 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=363063:\end{array}$
translation_(10.58)
$\mid$ | | | | | | | | | | start > 363063: meiosis_
(17.42)

conjugation_(11.86) | | | | | |
organelle_organization_and_biogenesis_ (15.68)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ start $>383302$

transport_(8.34/2.15)

biological_process_(9.26/2.06)
| | | | | | | | | | | distance > 1876:
transport_(8.26/3.13)
| | | | | | | | | start > 388821:
RNA_metabolic_process_(44.29)
$\begin{array}{lllllllll}\mid & - & \mid & \mid & \mid & \mid & \mid & \text { start }>392754 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=397902 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \quad \text { start }<=397076:\end{array}$
sporulation_(22.48)
| | | | | | | | | | start > 397076: translation_
(22.81)




| | | | | | | start <= 583920:
membrane_organization_and_biogenesis_(19.6)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | star $\bar{t}>583920$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=586387:$ |

organelle_organization_and_biogenesis_ (21.34)
| | | | | | | | start > 586387: translation_ (19.55)
 (20.33)
| | | | | | | | start > 594885:
cytoskeleton_organization_and_biogenesis_ (15.71)
| | | $|\quad| \quad \mid \quad$ strand $=W$
$|||||\mid$ neigh_num <=1: meiosis_(18.87/8.71)
$\left|\left|\left|\left|\left|\left|\mid n e i g h \_n u m>1:\right.\right.\right.\right.\right.\right.$ DNA_metabolic_process_
(21.7/12.26)





 biological_process_ (13.33)
| | | | | | | | | | | | start > 759806:
ribosome_biogenesis_and_assembly_ (34.94/16.74)
$|\quad| \quad|\quad| \quad-|\quad-|\quad| \quad \overline{\mid}| \quad$ strand $=\mathrm{W}$ | | | | | | | | | | start $<=$ 751960: organelle_organization_and_biogenesis_ (16.48)

| | | | | | | | | | | | | | | start <= 782030
| | | | | | | | | | | | | | | |
start <= 778000

| | | | | | | $\quad$ | | | | | | | | |
start > 770800: biological_process_ (30.08)
| | | | | | | | | | | | | | | |
start > 778000

start <= 779126: electron_transport_ (10.34)
| | | | | | | ${ }^{-}$| $\quad$ | | | | | | | |
start > 779126: translation_ (15.95)
| | | | | | | | | | | | | | | start >
782030: biological_process_(32.71)
| | | | | | | | | | | | | | start > 788882: translation (16.58)
| | | | | | | | | | | | | start > 789838:
response_to_stress_ (13.14)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ | $|\quad| \quad \mid \quad$ strand $=W$

| | | | | | | | | | | | | | |
769282: biological_process_(11.82)
| | | | | | | | | | | | | | start >
769282

| | | | | | | | | | | | | | |
start <= 770800: response_to_stress_ (15.82)
| | | | | | | | | $\mid$ | | | | | |
start > 770800: biological_process_ (10.74)



| | | | start > 904823: biological_process_(36.77)

## Decision Tree Generated for Chromosome Fourteen:

## For Cellular Component

start $<=155101$
$\mid$
$\mid$
$\mid$
$\mid$




```
| | | | | | | | | | NeighGO_term =
membrane fraction_: membrane fraction_(0.0)
| | \ | | | | \ | | NeighGO_term =
protein_binding_: site_of_polarized_growth_(3.19/1.76)
| | | | | | | | | NeighGO_term =
signal_transduction_: membrane_fraction_ (0.0)
| | | | | - | | | - | | - NeighGO_term = ribosome_:
membrane_fraction_(0.0)
| | \ | | | | | | | NeighGO_term = peroxisome_:
membrane_fraction_(0.0)
| | | | | | | | | | NeighGO_term =
oxidoreductase_activity_: site_of_polarized_growth_ (6.97/3.54)
| | | | | | | | | | NeighGO_term =
carbohydrate_metabolic_process_: membrane_fraction_(0.0)
| | | | | | | | | | NeighGO_term =
endoplasmic_reticulum_: cytoplasm_ (4.81/2.86)
| | | | | | | | | | NeighGO_term = membrane_:
membrane fraction (0.0)
| | | | | | | | | | NeighGO_term =
protein_catabolic_process_: membrane_fraction_(0.0)
| | | | | | | | | NeighGO_term =
translation_regulator_activity_: membrane_fraction_(0.0)
```



```
enzyme_regulator_activity_: membrane_fraction_(0.0)
| | | | | | | | | | NeighGO_term =
membrane_organization_and_biogenesis_: membrane_fraction_ (0.0)
| | | | | | | | | | NeighGO_term =
organelle_organization_and_biogenesis_: membrane_fraction_(0.0)
| | | | | | - | - | | | NeighGO_term =
DNA_metabolic_process_: membrane_fraction_(0.0)
| | | | | | | | | NeighGO_term =
protein_modification_process_: membrane_fraction_(0.0)
| | | | | | | | | | NeighGO_term =
transcription_: membrane_fraction_(0.0)
| | | | | | \ | | | | NeighGO_term = cell_cortex_:
membrane_fraction_(0.0)
| | | | | | | | | NeighGO_term =
cytoskeleton_: membrane_fraction_(0.0)
| | | | | - | \ | NeighGO_term =
protein_kinase_activity_: membrane_fraction_ (0.0)
| | | | | | | | | | NeighGO_term = cytokinesis_:
membrane_fraction_(0.0)
| | | | | | | | | | NeighGO_term = DNA_binding_:
membrane_fraction_(0.0)
| | \ | | | | | | | NeighGO_term =
transcription_regulator_activity_: membrane_fraction_ (0.0)
| | | | | | | | | NeighGO_term =
transferase_activity_: membrane_fraction_(0.0)
| | | | | | | | | | NeighGO_term =
Golgi_apparatus_: membrane_fraction_ (0.0)
| | | | | | | | | | NeighGO_term =
cell_wall_organization_and_biogenesis_: membrane_fraction_ (0.0)
| | | | | | | | | | NeighGO_term = chromosome_:
membrane_fraction_(0.0)
```








```
| | | | | | | | | | | | | | NeighGO_term
= structural_molecule_activity_: cellular_bud_ (3.48/0.97)
| | | | | | | | | | | | | | NeighGO_term
= endomembrane_system_: cellular_bud_(2.98/0.44)
| | | | | | | | | | | | | | NeighGO_term
= membrane_fraction_: cytoplasm_ (0.0)
| | | | | - | | | - | | | | | | NeighGO term
= protein_binding_: nucleus_ (1.92/1.09)
| | | | | | | | | | | | | | NeighGO_term
= signal_transduction_: cytoplasm_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= ribosome_: cellular_bud_ (3.53/1.75)
| | | | | | | | | | | | | | NeighGO_term
= peroxisome_: cytoplasm_(0.0)
| | | | | | | | | | | | | | NeighGO_term
= oxidoreductase_activity_: cytoplasm_ (0.0)
| | | | | | | | | | | | | | NeighGo_term
= carbohydrate_metabolic_process_: cytoplasm_(0.0)
| | | | | | | \ | \ | | \ | | NeighGO_term
= endoplasmic_reticulum_: nucleus_(2.63/0.08)
| | | | | | | | | | | | | | NeighGO_term
= membrane_: nucleus_(1.78/0.86)
| | | | | ` | | | | | | | | NeighGO_term
= protein_catabolic_process_: cellular_bud_ (1.3/0.67)
| | | | | | | | | | | | | | NeighGO_term
= translation_regulator_activity_: cytoplasm_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= enzyme_regulator_activity_: cytoplasm_ (0.0)
| | \ | | | | - | | | - | | | | NeighGO_term
= membrane_organization_and_biogenesis_: cytoplasm_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= organelle_organization_and_biogenesis_: cytoplasm_ (0.0)
| | | | | | | | | | | | | | N NeighGO_term
= DNA_metabolic_process_: cytoplasm_(0.0)
| | | | | | | | | | | | | | NeighGO_term
= protein_modification_process_: cytoplasm_ (0.0) | | | | | | | | | | | | | | | | NeighGO_term
= transcription_: cytoplasm_ (0.0)
| | | | - | | | - | | | | | | | NeighGO_term
= cell_cortex_: cytoplasm_(0.0)
| | | | | | | | | | | | | | NeighGO_term
= cytoskeleton_: cytoplasm_(0.0)
| | | | | | | | | | | | | | NeighGO_term
= protein_kinase_activity_: cytoplasm_( (0.0)
= cytokinesis_: cytoplasm_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= DNA_binding_: cytoplasm_(0.0)
| | | | | | | | | | | | | | NeighGO_term
= transcription_regulator_activity_: cytoplasm_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= transferase_activity_: cytoplasm_(1.97/0.91)
= Golgi_apparatus_: cytoplasm_ (0.0)
```





```
| | | | | | | | | | | | | | start <=
462413
| | | | | | | | | |
start <= 457706: nucleus_(53.44/22.67)
| | | | | | | | | | | | | | | |
start > 457706: cellular_bud_ (22.88)
| | | | | | \ \ | |
462413
| | | | | | | | | | | | | | | | |
start <= 467133: cytoplasm_ (45.23)
| | | | | | | | | | | | | | | |
start > 467133
| | | | | | | | | | | | | | | | |
start <= 471379: mitochondrion_ (31.59)
| | | | | |
| | | | | | | | | | | | | | | | |
| start <= 479769: nucleus (87.42/24.91)
| | | | | | | | | | | | | | | | |
| start > 479769: mitochondrion_ (82.25/46.2)
| | | | | | | | | | | | | | start >
4 8 1 3 9 2
| | | | | | | | | | | | | | start <=
485608
| | | | | | | | | | | | | | | | |
start <= 483557: cellular_component_ (38.5)
| | | | | | | | | | | | | | | | | |
start > 483557: nucleolus (21.25)
| | | | | | | | | | | | | | | start >
485608
| | | | | | | | | | | | | | | | | |
start <= 495002
| | | | | | | | | | | | | | | |
start <= 490318
| | | | | | | | | | | | | | | | |
| start <= 488387: mitochondrion_ (29.9)
```



```
| | | | | | | - | | |
start > 490318: mitochondrion (41.37)
| | | | | | | | | | | | | | | |
start > 495002: cellular_component_ (24.11)
| | | | | | | | | distance > 19559
| | | | | | | | | | start <= 444317: cytoplasm_
(65.0)
| | | | | | | | | | start > 444317
| | | | | | | | | | start <= 447613:
cytoskeleton_(4.0)
| | | \ | | | | | | | start > 447613:
cytoplasm_(16.62)
| | | | | | | | start > 499418
| | | | | | | | | start <= 503725
| | | | | | | | | | start <= 499682: ribosome_
(30.75)
```


| | | | | | | | | | | | | | start > 308614: plasma_membrane_ (17.55)
| | | | | | | | | | | | start > 310638: mitochondrion_(48.39)

cytoplasm_(28.75)
$\begin{array}{lllllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }>332599 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\ \text { NeighGO_term }=\end{array}$ nucleolus_: endomembrane_system_(0.0)
 RNA_binding_: endomembrane_system_ (0.0)
| $\left.{ }^{-}\left|{ }^{-}\right| \quad|\quad|\right|^{-}| |^{-}|\quad| \quad$ NeighGO_term $=$ RNA_metabolic_process_: endomembrane_system_ (0.0)
$\left|{ }^{-}\right| \quad|\quad| \quad|\quad| \quad|\quad| \quad{ }^{-}|\quad| \quad$ NeighGO_term $=$
ribosome_biogenesis_and_assembly_: nucleus_ (5.56/3.06)
 nucleus
| | | | | | | | | | | | distance <= 3429: nucleus_(6.45/2.61)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ distance $>$ 3429: endomembrane_system_(4.93/1.31)

 hydrolase_activity_: nucleus_(4.02/1.66)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGO_term $=$ cytoplasm_: membrane_(11.4/6.27)
| | | | $\mid$ | | | | | | | NeighGO_term $=$ molecular_function
| | | $\left.{ }^{-}\right|^{-}$| | | | | | | | | neigh_strand
= $W$ : nucleus_(4.29/1.51)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ neigh_strand $=C:$ endomembrane_system_(3.72/0.86)
 translation_: endomembrane_system_(0.0)
 meiosis_: endomembrane_system_(0.0)
$\left.|\quad| \quad|\quad| \quad|\quad| \quad\right|^{-}|\quad| \quad|\quad| \quad \mid \quad$ NeighGO_term $=$ pseudohyphal_growth_: endomembrane_system_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ | $\mid$ NeighGO_term $=$ biological_process_: endomembrane_system_(6.25/3.0) $\left.|\quad|\right|^{-}| |^{-}|\quad| \quad| |^{-} \mid \quad$ | $\mid$ NeighGO_term $=$ transport_: endomembrane_system_(1.28/0.59)


| | | | | | | | | | | microtubule_organizing_center_: endomembrane_system_(0.0) | | | | | | | | | | | | | NeighGO_term = sporulation_: endomembrane_system_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGO_term $=$ vitamin_metabolic_process_: endomembrane_system_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|-|\quad| \quad| \quad$ | $\mid$ NeighGO term $=$ extracellular_region_: endomembrane_system_(0.0)
 nuclear_organization_and_biogenesis_: endomembrane_system_(0.0)
$|\quad| \quad-|\quad| \quad|\quad \overline{\mid} \quad \overline{\mid}| \quad-|\quad| \quad|\quad|-$ Neigh $\bar{G} O \_t e r m=$ cell_budding_: endomembrane_system_(0.0)
| $\mid$ | | | | - | $\mid$ | | | mitochondrial_envelope_: endomembrane_system_(0.0)
 cytoplasmic_membrane-bound_vesicle_: endomembrane_system_( $\overline{0} .0$ )
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ start ${ }^{-}>3403 \overline{5} 4$
| | | | | | | | | | 1 start $<=345278$ | | | | | | | | | | | | start <= 342518: endoplasmic_reticulum_ (32.5)
plasma_membrane_(27.55)
 cytoplasm_(22.51)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ start $>346315$ 349758
| | | | | | | | | | | | | | start $<=$ 348523: nucleus_(25.68)
| | | | | | | | | | | | | | | start > 348523: endoplasmic_reticulum_(20.55)
| | | | | | | | | | | | | | start >
349758: nucleus_(50.07/19.95)

366743
| | | | | | | | | | | | | | | |
start $<=$ 359598: mitochondrion (18.3)
| | | | | | | | | | | | | | | |
start > 359598: nucleus_ (27.74)
| | | | | | | | | | | | | | | start >
366743: mitochondrion_(32.0)
| | | | | $\left.\right|^{-} \mid$| | | | | 1 start $>370370$
373583

```
| | | | | | | | | | | | | | | start <=
372228
```



```
    NeighGO_term = RNA_binding_: nucleus__(0.0)
    | | - | | - | | | | | \ | | | | | | |
NeighGO_term = RNA_metabolic_process_: nucleus_ (0.0)
| | | | | | | | | | | | | | | | |
NeighGO_term = ribosome_biogenesis_and_assembly_: nucleus_
(1.8/0.\overline{6}3)
| | | | | | | | | | | | | | | |
NeighGO_term = nucleus_
neigh_strand = W: cytoplasm_(2.98/0.58)
neigh_strand = C: nucleus_ (11.67/7.2)
| |
NeighGO_term = cellular_component_: nucleus_ (0.0)
```



```
| | | | | | | | | | | | | | | | |
NeighGO_term = cytoplasm_: mitochondrion_ (5.57/2.62)
| | | | | | | | | | | | | | | |
NeighGO_term = molecular_function_: mitochondrion_(11.63/6.49)
NeighGO_term = translation_: mitochondrion_(3.65/1.66)
| | | | | | | | | | | | | | | | | |
NeighGo_term = meiosis_: nucleus_ (0.0)
| | | | | | | | | | | | | | | |
NeighGO_term = pseudohyphal_growth_: nucleus_ (4.73/1.99)
NeighGO_term = biological_process_: nucleus_(1.4/0.71)
| | - | | | | | | | | | | | - | | | | |
NeighGO_term = transport_: nucleus_(9.43/4.01)
| | | | | | | | | | | | | | | | |
NeighGO_term = mitochondrion_: cytoplasm_ (5.21/2.32)
| | | | | | | | | | | | | | | |
NeighGO_term = transporter_activity_: mitochondrion_(1.43/0.26)
| | | | | | | | | - | | | | - | | |
NeighGO_term = structural_molecule_activity_: nucleus_(1.34/0.73)
| | | | | | | | | | | | l | | | | | |
NeighGO_term = endomembrane_system_: nucleus_ (0.0)
| | - | | | | - | | - | | \ \ | | | |
NeighGO_term = membrane_fraction_: nucleus_ (0.0)
NeighGO_term = protein_binding_: cytoplasm_(3.24/1.55)
NeighGO_term = signal_transduction_: nucleus_ (0.0)
| | - | | | | | | | - | | \ \ | | | |
NeighGO_term = ribosome_: nucleus_ (0.0)
NeighGO_term = peroxisome_: nucleus_(0.0)
```

$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad|$ NeighGO_term = oxidoreductase_activity_: nucleus_(3.52/0.79) | | | | | | | | | | | | | | NeighGo_term = carbohydrate_metabolic_process_: nucleus_(0.0)

 NeighGO_term = membrane_: nucleus_ (0.0)
| | | | | | | | | | | | | | | | NeighGO_term = protein_catabolic_process_: nucleus_ (0.0) | | | | | | | | | | | | | | | | NeighGO_term = translation_regulator_activity_: nucleus_ (0.0) | | | | | | | NeighGO_term = enzyme_regulator_activity_: nucleus_ (0.0)

NeighGO_term = membrane_organization_and_biogenesis_: nucleus_ (0.0)
NeighGo_term = organelle_organization_and_biogenesis_:
mitochondrion_ (4.52/2.17)
$\left\lvert\, \begin{array}{cc}\mid & \mid \\ \text { NeighGo_term } & \mid \\ \mid & \mid \\ \text { DNA_metabolic_process }\end{array}\right.$
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad 1$
neigh_num $<=1$ : nucleus_(4.57/2.33)
| | | | | | - | | |
neigh_num > 1: cytoplasm_(5.87/2.0)
| | | | | | | | | | | | | | |
NeighGo_term = protein_modification_process_: nucleus_ (0.0)
| | | | | | | | | | | | | | | |
NeighGO_term = transcription_: nucleus_ (0.0)
NeighGO_term = cell_cortex_: cytoplasm_(7.3/1.76)
| | | | | | | | | | | | | | | |
NeighGO_term = cytoskeleton_: mitochondrion_(2.66/0.97)
| | - | | | | - | | | |
NeighGo_term = protein_kinase_activity_: nucleus_ (0.0)
| | | | | | | | | | | | | | | |
NeighGO_term = cytokinesis_: nucleus_ (0.0)
| | | | | | | | | | | | | | | |
NeighGO_term = DNA_binding_: mitochondrion_(1.71/0.98)
| | | | | | | | | | | ${ }^{-}$| | | |
NeighGo_term = transcription_regulator_activity_: nucleus_ (0.0)

NeighGO_term = transferase_activity_: nucleus_ (0.0)

NeighGo_term = Golgi_apparatus_: nucleus_(0.0)
| | | | | | | | | | | | | | | |
NeighGO_term = cell_wall_organization_and_biogenesis_: nucleus_
(0.0)
NeighGo_term = chromosome_: nucleus_(0.0)

NeighGO_term = phosphoprotein_phosphatase_activity_: nucleus_ (0.0)

NeighGo_term = response_to_stress_: nucleus_ (0.0)
| | | | | | | | | | | | | | | | NeighGo_term = lyase_activity_: nucleus_ (0.0) NeighGO_term = cellular_respiration_: nucleus_ (0.0)

NeighGO_term = vesicle-mediated_transport_: nucleus_(0.0)

NeighGO_term = peptidase_activity_: nucleus_ (0.0)
| | | | | | | | | | | | | | | |
NeighGo_term = lipid_metabolic_process_: nucleus_ (0.0)


NeighGo_term = cytoskeleton_organization_and_biogenesis_: nucleus_
(2.62/1.1)
NeighGo_term = electron_transport_: nucleus_(0.0)
| | ${ }^{-}|\quad| \quad\left|{ }^{-}\right| \quad|\quad|-\quad|\quad|{ }^{-}|\quad| \quad|\quad| \quad \mid$
NeighGo_term = conjugation_: nucleus_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad|$
NeighGo_term = vacuole_: nucleus_(0.0)
NeighGO_term = helicase_activity_: nucleus_ (0.0)
NeighGO_term = cell_wall_: nucleus_ (0.0)
NeighGo_term = ligase_activity_: nucleus_(0.0)
$\left.|\quad| \quad|\quad| \quad|\quad|-|\quad|-\quad|\quad| \quad \begin{aligned} & - \\ & \mid\end{aligned} \right\rvert\,$
NeighGo_term = cellular_bud_: nucleus_(0.0)
| | | | | | | | | |
$\underset{\mid}{\text { NeighGO_term }=~ s i t e ~} \mid$
NeighGO_term = cell_cycle_: nucleus_(0.0)

NeighGO_term = isomerase_activity_: nucleus_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad|$
NeighGO_term = signal_transducer_activity_: nucleus_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad|$
NeighGo_term = amino_acid_and_derivative_metabolic_process_:
nucleus_ (0.0)
| | | | | | | | | | | | | | |
NeighGO_term = nucleotidyltransferase_activity_: nucleus_ (0.0)
NeighGO term = microtubule organizing center : nucleus (0.0)

NeighGo_term = sporulation_: nucleus_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|-|\quad| \quad| \quad|\quad| \quad|\quad| \quad|\quad|$
NeighGO_term = vitamin_metabolic_process_: nucleus_ (0.0)
| | | | | | | | | | | | | | | |
NeighGO_term = extracellular_region_: nucleus_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad-\quad| \quad|-|$
NeighGO_term = nuclear_organization_and_biogenesis_: nucleus_ (0.0)
| | | | | | | 1 | 1
NeighGo_term = cell_budding_: nucleus_(0.0)
| | | | | | | | | | | | | | | | NeighGO_term = mitochondrial_envelope_: nucleus_(0.0) $|\quad|-|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad-|\quad| \quad|\quad|$ NeighGO_term = cytoplasmic_membrane-bound_vesicle_: nucleus_(0.0) | | | | | | | | | | | | | | | start > 372228
| | | | | | | | | | | | | | | | neigh_strand $=W$ W: cytoplasm_(35.57/14.08)
| | ${ }^{-}$| | | | neigh_strand = C: nucleus_(26.59/11.89)
| | | | | | | | | | | | | | start > 373583: nucleus_ (34.25)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>375323$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| $\mid$ | start $<=383989$ |  |  |  |  |  |  |  |  |  |  |  |
| start $<=$ |  |  |  |  |  |  |  |  |  |  |  |  |

380833: endoplasmic_reticulum_(23.52)
| | | | | | | | | | | | | start > 380833: cellular component_(34.96)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ start $>383989$ 401042
| | | | | | | | | | | | | | | start $<=$ 394687
| | | | | | | | | | | | | | |
neigh_num <= 1: mitochondrion_(36.86/14.52)

neigh_num > 1: membrane_(22.9/9.58)
| | ${ }^{-}$| | | - | | | | | | | | start > 394687: mitochondrion (51.31)
| | | | | | | | | | | | | | start > 401042: nucleus_(52.34/22.52)
$\begin{array}{llllllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }>408343 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=417304\end{array}$ $||||||||||\mid$ start $|=$ 413641: vacuole_(20.97)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ start $>413641$ $\mid$ | | | | | | | | | | | | NeighGO_term = nucleolus
| | | $-\mid$ | | | | | | | | | neigh strand $=W$ : endomembrane system (2.08/0.99)
| | ${ }^{-}$| | | | | ${ }^{-}$| $\left.\right|^{-}$| | | |
neigh_strand = C: membrane_fraction_(8.55/3.05)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad N e i g h G O \_t e r m$ = RNA_binding_: endoplasmic_reticulum_ (0.0)
| | ${ }^{-}$| ${ }^{-}$| | $\mid$| | = RNA_metabolic_process_: endoplasmic_reticulum_(0.0)
| | | | | | | | | | | | | | NeighGO_term = ribosome_biogenesis_and_assembly_

neigh_strand $=W:$ membrane_ (4.16/2.01)
| | | | | | | | | | | | | |
neigh_strand $=C$ : endomembrane_system_(4.42/2.06)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGO_term
= nucleus


```
| | | | | | | | | | | | | | NeighGO_term
= translation_regulator_activity_: endoplasmic_reticulum_(0.0)
| | | | | | | | | | | | | | NeighGO_term
= enzyme_regulator_activity_: endoplasmic_reticulum_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= membrane_organization_and_biogenesis_: endoplasmic_reticulum_
(0.0)
| | | | | | | | | | | | | | NeighGO_term
= organelle_organization_and_biogenesis_: endoplasmic_reticulum_
(0.0)
| | | | | | | | | | | | | | NeighGO_term
= DNA_metabolic_process_: endomembrane_system_(10.2/6.98)
| | | | | | | | | | | | | | NeighGO_term
= protein_modification_process_: endoplasmic_reticulum_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= transcription_: endoplasmic_reticulum_ (0.0)
| | | | | | | | | | - | | | | NeighGO_term
= cell_cortex_: endoplasmic_reticulum_(0.0)
| | | | | | | | | | | | | | NeighGO_term
= cytoskeleton_: endoplasmic_reticulum_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= protein_kinase_activity_: endoplasmic_reticulum_ (0.0)
| | | | ` | | ` | | | - | | | | | NeighGO_term
= cytokinesis_: endoplasmic_reticulum_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= DNA_binding_: endoplasmic_reticulum_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= transcription_regulator_activity_: endoplasmic_reticulum_ (0.\overline{0})
| | | | - | | | | | | | | \ | NeíghGO_term
= transferase_activity_: endoplasmic_reticulum_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= Golgi_apparatus_: endoplasmic_reticulum_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= cell_wall_organization_and_biogenesis_: endoplasmic_reticulum_
(0.0)
| | | | | | | | | | | | | | NeighGO_term
= chromosome_: endoplasmic_reticulum_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= phosphoprotein_phosphatase_activity_: endoplasmic_reticulum (
| | | | | | | | | | | | | | NeighGO_term
= response_to_stress_: endoplasmic_reticulum_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= lyase_activity_: endoplasmic_reticulum_ (0.0)
| | | | | | | | | | \ | | | NeighGO_term
= cellular_respiration_: endoplasmic_reticulum_(0.0)
| | | | | | | | | | | | | | NeighGO_term
= vesicle-mediated_transport_: endoplasmic_reticulum_(0.0)
| | | | | | | | | | | | | | NeighGO_term
= peptidase_activity_: endoplasmic_reticulum_ (0.0)
| | | - | | | | | | | | | \ | | NeighGO term
= lipid_metabolic_process_: endoplasmic_reticulum_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= plasma_membrane_: endoplasmic_reticulum_ (0.0)
```

| | | | | | | | | | | | | | NeighGO_term = cytoskeleton_organization_and_biogenesis_: endoplasmic_reticulum_ (0.0)
| | | | | | | | | | | | | | NeighGO_term = electron_transport_: endoplasmic_reticulum_ (0.0)
$\left.|\quad|\right|^{-}|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGo_term = conjugation_: endoplasmic_reticulum_(0.0)

= vacuole_: membrane_fraction_(3.8/2.2)
$\left.|\quad|\right|^{-}|\quad| \quad{ }^{-} \quad| |^{-}|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGO_term = helicase_activity_: endoplasmic_reticulum_(3.81/1.33)
$\left.\left.|\quad|\right|^{-}\left|{ }^{-}\right| \quad|\quad|\right|^{-}|\quad|{ }^{-}|\quad| \quad \mid \quad$ NeighGO_term = cell_wall_: endoplasmic_reticulum_ (0.0)
| | | | | | | | | | | | | | NeighGO_term
= ligase_activity_: endoplasmic_reticulum_(0.0)
 = cellular_bud_: endoplasmic_reticulum_ (0.0)
| | | | ${ }^{-}$| | | | | ${ }^{-}$| | | NeighGO_term = site_of_polarized_growth_: endoplasmic_reticulum_ (0.0)
$\left.\left|\left.\right|^{-}\right|^{-}|\quad|\right|^{-}|\quad| \quad|\quad| \quad|\quad|-\quad \mid \quad$ NeighGO_term = cell_cycle_: endoplasmic_reticulum_(0.0)
| | | | | | | | | | | | | | NeighGo_term = isomerase_activity_: endoplasmic_reticulum_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGo_term = signal_transducer_activity_: endoplasmic_reticulum_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ | $\mid$ | $\mid$ | $\mid$ |
= amino_acid_and_derivative_metabolic_process_:
endoplasmic_reticulum_(0.0)
| | | | | | | | | | | | | NeighGO_term = nucleotidyltransferase_activity_: endoplasmic_reticulum_
(6.21/1.82)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGO_term = microtubule_organizing_center_: endoplasmic_reticulum_ (0.0)
 = sporulation_: endoplasmic_reticulum_(0.0)
| | | | | | | | | | | | | | NeighGO_term = vitamin_metabolic_process_: endoplasmic_reticulum_ (0.0)
| | | | | | | | | | | | | | NeighGO_term = extracellular_region_: endoplasmic_reticulum_ (0.0)
| | | | | | | | | | | | | | NeighGO_term = nuclear_organization_and_biogenesis_: endoplasmic_reticulum (0.0)

= cell_budding_: endoplasmic_reticulum_ (0.0)
| | | | ${ }^{-}$| | $\mid$| $\left.\right|^{-}$| | | = mitochondrial_envelope_: endoplasmic_reticulum_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ | $\mid$ NeighGO_term
= cytoplasmic_membrane-bound_vesicle_: endoplasmic_reticulum_( $\overline{0} .0$ )
| | | | | | | | | | | 1 start > 417304

420100: nucleus_(111.91/50.28)
| | | | | | | | | | | | | | start >
420100


| | | | | | | | | | | NeighGO_term = endoplasmic_reticulum_: membrane_fraction_ (0.0)
| | | | | | | | $\mid$ | | | NeighGo_term = membrane_: membrane_fraction_(0.0) | | | | | | | protein_catabolic_process_: membrane_fraction_ (0.0)
$|\quad| \quad|\quad| \quad|\quad|-\left.\quad|\quad| \quad\right|^{-}$NeighGoterm = translation_regulator_activity_: membrane_fraction_(0.0)
| | | | | | | | | | 1 NeighGo_term = enzyme_regulator_activity_: membrane_fraction_(0.0)
| | | | | | | | | | | NeighGO_term = membrane_organization_and_biogenesis_ membrane_fraction (0.0) | | $\mid$ | | | | | | $\mid$ | NeighGo_term = organelle_organization_and_biogenesis_: membrane_fraction_ (2.22/0.41)
| | | | | | | | | | | 1 NeighGO_term = DNA_metabolic_process_: membrane_fraction_(0.0)
| $-\quad| |^{-}| |^{-}|\quad| \quad|\quad|-\quad$ NeighGo_term $=$
protein_modification_process_: membrane_fraction_(0.0)
| | | | | | | | | | | NeighGo_term =
transcription_: membrane_fraction_ (0.0)

cell_cortex_: membrane_fraction_ (0.0)

cytoskeleton_: membrane_fraction_(0.0)
| | | | | | | | | |
protein_kinase_activity_: membrane_fraction_ (0.0)
| | | | | | | | | | | NeighGO_term =
cytokinesis_: membrane_fraction_(0.0)
| | | | | | | | | | | | NeighGo_term =
DNA_binding_: cytoplasm_(1.46/0.19) , Neichen term =
transcription_regulator_activity_: membrane_fraction_( $\overline{0} .0$ )
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ | $\mid$ Neigh̄Go_term $=$
transferase_activity_: membrane_fraction_(5.14/1.16)
| | | | | | | | | | | | NeighGO_term =
Golgi_apparatus_: membrane_fraction_(0.0)
| | | | | | | | | | | | NeighGO_term =
cell_wall_organization_and_biogenesis_: membrane_fractiōn_(0.0)
| $\quad|-|\quad| \quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \bar{N} e i g h G o \_t e ̄ r m=$
chromosome_: membrane_fraction_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGO_term $=$
phosphoprotein_phosphatase_activity_: membrane_fraction_(0.0)
| | | | | | | | | | | | NeighGO_term =
response_to_stress_: membrane_fraction_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ | $\mid$ NeighGo_term $=$
lyase_activity_: membrane_fraction_(0.0)
cellular_respiration_: membrane_fraction_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad$ | $\mid \quad$ NeighGo_term $=$
vesicle-mediated_transport_: membrane_fraction_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGo_term $=$
peptidase_activity_: membrane_fraction_ (0.0)




```
| | | | | | | | | | | | | start > 586821:
cytoplasm_(24.19)
| | | | | | | | | | | | start > 587848:
cellular_component_ (32.01)
| | | | | | | | | start > 591161: nucleus_
(36.48)
| | | | | | | | | start > 593228
| | | | | | | | | start <= 597540:
cell_cortex_(58.46/26.31)
| | | | | | | | | | start > 597540:
cellular_component_ (96.91)
| | \ | | | distance > 20940
| | | | | | | start <= 576728: cytoplasm_ (99.44)
| | | | | | | start > 576728
```



```
cellular_component_ (13.25)
| | | | | | | | | start > 585588: nucleolus_
(28.18)
| | | | | | | | start > 593228: cytoplasm_ (21.36)
| | | start > 614822
| | | | start <= 620069
| | | | | start <= 618221
| | | | | | neigh_strand = W
| | | | | | | NeighGO_term = nucleolus_:
endomembrane_system_(0.0)
| | | | | | | NeighGO_term = RNA_binding_:
endomembrane_system_(10.51/2.44)
| | | \ | - | | NeighGO_term = RNA_metabolic_process_:
endomembrane_system_(0.0)
| | | | | - | | NeighGO_term =
ribosome_biogenesis_and_assembly_: endomembrane_system_(0.0)
| | | | | | | NeighGO_term = nucleus_:
endomembrane system (9.03/3.83)
| | | | | | | NeighGO_term = cellular_component_:
endomembrane_system_ (0.0)
| | | | | | | NeighGO_term = hydrolase_activity_:
endomembrane_system_(0.0)
| | | | | | | NeighGO_term = cytoplasm_:
endoplasmic_reticulum_(8.92/2.9)
| | | | | | NeighGO_term = molecular_function_:
endoplasmic_reticulum_(2.63/0.83)
| | | | | | NeighGO_term = translation_:
endomembrane_system_(0.0)
| | | \ | - | | NeighGO_term = meiosis_:
endomembrane_system_ (0.0)
| | | | | | NeighGO_term = pseudohyphal_growth_:
endomembrane_system_(0.0)
| | | | | | | NeighGO_term = biological_process_:
endoplasmic_reticulum_(11.45/6.01)
| | | | | | | NeighGO_term = transport_:
endoplasmic_reticulum_ (5.78/1.32)
| | | | | | | NeighGO_term = mitochondrion_:
endomembrane_system_(3.2/1.67)
```

```
| | | | | | NeighGO term = transporter activity :
endomembrane_system_(0.0)
| | | | | | | NeighGO_term =
structural_molecule_activity_: endomembrane_system_(0.0)
| | | | | | | NeighGO_term = endomembrane_system_:
endoplasmic_reticulum_(1.33/0.49)
| | - | | | | NeighGO_term = membrane_fraction_:
endomembrane_system_(4.6/1.54)
| | | | | | | NeighGO_term = protein_binding_:
endoplasmic_reticulum_(1.65/0.55)
| | | | | | | NeighGO_term = signal_transduction_:
endoplasmic_reticulum_(6.15/3.27)
| | | | | | | NeighGO_term = ribosome_:
endomembrane_system_(0.0)
| | | | | | | NeighGO_term = peroxisome_:
endomembrane_system_(3.94/1.93)
| | | | | | NeighGO_term = oxidoreductase_activity_:
membrane (1.45/0.49)
| | \ | | | | NeighGO_term =
carbohydrate_metabolic_process_: membrane_(3.22/1.62)
| | | | | | | NeighGO_term = endoplasmic_reticulum_:
endomembrane_system_(0.0)
| | | | | - | | NeighGO_term = membrane_:
endomembrane_system_(0.0)
| | | | | | | NeighGO_term =
protein_catabolic_process_: endomembrane_system_ (0.0)
| | | | | | NeighGO_term =
translation_regulator_activity_: endomembrane_system_(0.0)
| | | - | | | NeighGO_term =
enzyme_regulator_activity_: endomembrane_system_ (0.0)
| | | | | | | NeighGO_term =
membrane_organization_and_biogenesis_: endomembrane_system_ (0.0)
| | | | | | | NeighGO term =
organelle_organization_and_biogenesis_: endomembrane_system_(0.0)
| | | | | | | NeighGO_term = DNA_metabolic_process_:
endomembrane_system_(0.0)
| | | | | | | NeighGO_term =
protein_modification_process_: endomembrane_system_(0.0)
| | - | | \ | N
endomembrane_system_(0.0)
| | | | | | | NeighGO_term = cell_cortex_:
endomembrane_system_ (0.0)
| | | | | | | NeighGO_term = cytoskeleton_:
endomembrane_system_(0.0)
| | | \ | - | | NeighGO_term = protein_kinase_activity_:
endomembrane_system_ (0.0)
| | | | | | | NeighGO_term = cytokinesis_:
endomembrane_system_ (0.0)
| | | | | | | NeighGO_term = DNA_binding_:
endomembrane_system_(0.0)
| | | | - | | NeighGO_term =
transcription_regulator_activity_: endomembrane_system_ (0.0)
| | | | | | | NeighGO_term = transferase_activity_:
endomembrane_system_(0.0)
```





$\mid$
$\mid$
$\mid$


```
| | | | | | | NeighGO term =
membrane_organization_and_biogenesis
| | | | | | | NeighGO_term =
organelle_organization_and_biogenesis_: nucleus_ (0.0)
| | | | | | | NeighGO_term = DNA_metabolic_process_:
nucleus_ (0.0)
| | | | | | | NeighGO term =
protein_modification_process_: nucleus_ (0.0)
| | | | | | | NeighGO_term = transcription_: nucleus_
(0.0)
| | | | | | | NeighGO_term = cell_cortex_: nucleus_
(0.0)
| | | | | | | NeighGO_term = cytoskeleton_: nucleus_
(0.0)
| | | | | | | NeighGO_term = protein_kinase_activity_:
nucleus_(0.0)
| | | | | | | NeighGO_term = cytokinesis_: nucleus_
(0.0)
| | | | | | | NeighGO_term = DNA_binding_: nucleus_
(0.0)
| | | | | | | NeighGO_term =
transcription_regulator_activity_: nucleus_ (0.0)
| | | | | | | Neigh̄GO_term = -transferase_activity_:
nucleus_(0.0)
| | | | | | | NeighGO_term = Golgi_apparatus_:
nucleus_(0.0)
| | | | | | | NeighGO_term =
cell_wall_organization_and_biogenesis_: nucleus_(0.0)
| \ | | | | - | - NeighGo_tērm = chromosome_: nucleus_
(0.0)
| | | | | | | NeighGO_term =
phosphoprotein_phosphatase_activity_: nucleus_(0.0)
| | | | | | | NeighGO_term = response_to_stress_:
nucleus_(0.0)
| | | | | | | NeighGO_term = lyase_activity_: nucleus_
(0.0)
| | | | | | | NeighGO_term = cellular_respiration_:
nucleus_ (0.0)
| | | | | | | NeighGO term = vesicle-
mediated_transport_: nucleus_(4.73/1.4)
| | | | | | NeighGO_term = peptidase_activity_:
nucleus_(0.0)
| | | | | | | NeighGO_term = lipid_metabolic_process_:
nucleus_(0.0)
| | | | | | | NeighGO_term = plasma_membrane_:
nucleus_(4.31/0.7)
| | | | | | NeighGO_term =
cytoskeleton_organization_and_biogenesis_: nucleus_ (0.0)
| | | | | | NeighGO_term = electron_transport_:
nucleus_(0.0)
| | | | | | | NeighGO_term = conjugation_: nucleus_
(0.0)
| | | | | | | NeighGO_term = vacuole_: nucleus_(0.0)
```



## For Molecular Function

start <= 703701

```
start <= 646952
    | start <= 87897
        | start <= 52661
            start <= 40328
            | start <= 22633
                        | | start <= 14832
```

```
| | | | | | | | start <= 6098: helicase_activity
(12.72)
| | | | | | | | | | | | l start > 6098 
molecular_function_(56.18)
| | | | | | | | | start > 12876
| | | | | | | | | start <= 13267:
protein_binding_ (35.13)
| | | | | | | | | | start > 13267:
molecular_function_(27.86)
| | | | | | | start > 14832
| | | | | | | start <= 17248:
oxidoreductase_activity_ (31.95)
| | | | | | | | start > 17248
| | | | | | | | | start <= 19302:
transcription_regulator_activity_ (30.21)
| | | | | | | | \ start > 19302: protein_binding_
(25.75)
\begin{tabular}{lllllll}
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & start \(>22633\) \\
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & start \(<=31943\) \\
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & start \(<=23274:\)
\end{tabular}
transporter_activity_ (63.81/23.69)
| | | - | \ | | start > 23274
| | | | | | | | start <= 28346
| | | | | | | | | start <= 24047:
hydrolase_activity_ (30.41)
| | | | | | | | | | start > 24047:
transferase_activity_ (22.01)
| | | | | ` | | | start > 28346:
hydrolase_activity_ (64.05)
| | | | | | | start > 31943
| | | | | | | start <= 34234:
structural_molecule_activity_ (21.46)
| | | | | - | | ` start > 34234
| | | | | | | | | strand = W
| | | | | | | | | | start <= 34695:
transporter_activity_ (28.08)
| | | | | | | | | | start > 34695:
molecular function (31.17)
| | | | | | | | | strand = C:
transporter_activity_ (29.3)
| | | | | start > 40328
| | | | | | strand = W
| | | | | | start <= 40618: RNA_binding_ (22.24)
| | | | | | start > 40618
| | | | | | | | start<= 44446:
transcription_regulator_activity_ (27.14)
| | | | | | | | start > 44446
| | | | | | | | | start <= 48286: DNA_binding_
(36.17)
| | | | | | | | | start > 48286:
transcription_regulator_activity_ (21.23)
| | | | | | strand = C
| | | | | | | start <= 43074: lyase activity (29.49)
```













| | | | | | | | | | | | start <= 349758: molecular function (24.45)
| | | | | | | | | | | | | start > 349758:
protein_binding_ (14.94)
| | | | | | | | | | | start > 350942.
molecular function (41.66)

| $\mid$ | $\mid$ | $\left.\right\|^{-}$ | $\mid$ | $\left.\right\|^{-}$ | $\mid$ | $\mid$ | $\mid$ | start $>359598$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| $\mid$ | start $<=370370$ |  |  |  |  |  |  |  |
| 1 | start $<=366743:$ |  |  |  |  |  |  |  |

DNA_binding_(26.7)
| | | | | | | | | | | | $\mid$ | $\mid$ start > 366743:
| | | | | - | $\mid$ | | start > 370370:
isomerase_activity_ (32.22)
| | | | | |
| | | | | | | | | start <= 420100
| | | | | | | | | | start <= 380833
| | | | | | | | | | start <= 373583:
oxidoreductase_activity_ (28.58)

molecular function (26.63)
| | | ${ }^{-}$| | | | | | | | ${ }^{-}$|
transferase_activity_ (19.88)
$\begin{array}{lllllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }>380833 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=401042\end{array}$
| | | | | | | | | | | | start <= 383989:
molecular_function_(33.41)
| | | | | | | | | | | | | start > 383989
| | | | | | | | | | | | | | start<=

394687: transporter_activity_ (27.26)
| | | | | | | | | | | | | | start > 394687
| | | | | | | | | | | | | | start $<=$

```
398372: molecular_function_(20.29)
```

| | | | | | | | | | | | | | | start >
398372: transporter activity (23.95)
| | | | | - | | | | | | start > 401042
| | | | | | | | | | $\mid$ start $<=408343$ :
hydrolase_activity_ (14.16)
| | | | | | | | | | | | | start > 408343
417304
| | | | | | | | | | | | | | | start $<=$
413641: molecular function (22.6)
| | | | | | | | | | | | | | | start >
413641: oxidoreductase_activity_ (24.36)
| | | | | | | | | | | | | | start >
417304: molecular_function (46.45)






## For Biological Process



```
| | | | | | | start <= 14832:
vitamin_metabolic_process_ (76.66)
| | | | | | | start > 14832: biological_process_
(17.68)
| | | | | start > 17248: protein_modification_process_
(37.76/16.31)
| | | | start > 19302
| | | | | start <= 31943
| | | | | start <= 24047
| | | | | | strand =W: cytokinesis_(36.52)
| | | | | | | strand = C
s start <= 22633: signal transduction
(29.37)
\begin{tabular}{lllllllll}
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & start \(>22633\) \\
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid \quad\) neigh_strand \(=W\)
\end{tabular}
organelle_organization_and_biogenesis_(15.63/3.91)
| | | | | | | | | neigh_strand = C: transport_
(36.7/13.93)
| | | | | | | | | | | | start > 24047 
membrane_organization_and_biogenesis_(67.92/23.83)
| | | | | | | start > \overline{28346}
| | | | | | | start <= 31377: conjugation_(28.48)
membrane_organization_and_biogenesis_ (37.99/18.36)
| | | | start > 31943 
| | | | | | start <= 34234:
cell_wall_organization_and_biogenesis_ (17.53)
| | | | | | | - | - start > 3\overline{4}234
| | | | | | | start <= 34695: transport_ (35.64)
| | | | | | | | start <= 37699:
biological_process_(25.54)
| | | | | | | | | start > 37699: transport_
(19.37)
| | | | | | start > 40328
| | | | | | | strand = W: RNA_metabolic_process_
(26.13)
| | | | | | | strand = C
| | | | | | | start <= 43074:
amino_acid_and_derivative_metabolic_process_ (32.86)
| | | | | | | | start > 43074:
organelle_organization_and_biogenesis_ (20.59)
| | | start > 44\overline{2}79
| | | | start <= 491524
| | | | | start <= 350673
start <= 350673
| strand = W
    | start <= 117349
    | | | start <= 96312
```



| | | | | | | | | | | | | | | start > 202429: biological process (24.0)
$|\quad| \quad|\quad|-\left|\left.\right|^{-}\right| \quad|\quad| \quad|\quad| \quad$ start $>205189$
211923
| | | | | | | | | | | | | | | start <= 206931: transcription (19.81) 206931
| | | | | | | | | | | | | | | |
start < $=210234$ : response_to_stress_ (21.74)
| | | | | | $\left.\right|^{-} \quad$ | $\quad$ - | | | | | | start > 210234: cytokinesis_ (14.7)
| | | | | | | | | | | | | | start >
211923
1
227371 $|$ | | | | | | | | | | | | start $<=$
| | | | | | | | | | | | | | | | start <= 220202: ribosome_biogenesis_and_assembly_ (16.87)

start > 220202: organelle_organization_and_biogenesis_ (29.82)
| | | | | | | | | | | | | | | start > 227371: ribosome_biogenesis_and_assembly_ (24.98)
| | | | $\rceil$ | | - | | | $\mid$ | start > 230673 | | | | | | | | | | | | | start <= 238239: DNA_metabolic_process_(69.66)

244469
| | | | | | | | | | | | | | | start <=
240332: biological_process_ (32.47)
240332

| | | | | | | | | | | | | | | |
start > 241690: biological_process_ (28.76)
| | | | | | | | | | | | | | start >
244469
| | | | | | | | | | | | | | | start <=
245618: transport_ (49.73/21.66)
| | | | | | | | | | | | | | | start >
245618: biological_process_ (24.09)
| | | | | | | | | | start > 247462
$\left\lvert\, \begin{array}{llllllllll}\text { | } & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=250932 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { neigh_strand }=W \text { : }\end{array}\right.$
DNA_metabolic_process_ (53.58/28.64)

cell_cycle_(34.32/17.5)
| ${ }^{\text {| }}{ }^{-}$| | | | | | | | start > 250932
| | | | | | | | | | | | start <= 252060:
translation_(32.38)
| | | | | | | | | | | | start <= 254419: biological_process_ (17.85)
| | | | | | | | | | | | | start > 254419:
ribosome_biogenesis_and_assembly_ (25.13)
| | $\quad|\quad| \quad|\quad-| \quad$ | $\quad$ strand $=C$
| | | | | | | start $<=160374$
$\begin{array}{lllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=60297 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\ \mid & \text { start }<=51687\end{array}$
| | | | | | | | | | | start <= 48286:
biological_process_ (21.85)
| | | | | | | | | | | start > 48286: protein_catabolic_process_ (21.14)


76946: transport_ (21.26)
| | | | | | | | | | | | | | start > 76946
| | | | | | | | | | | | | | start $<=$ 82806: sporulation_(59.65/23.97)
$\mid$ | | | | | | | | | | | | | | start >
82806: transport_ (20.97)
| | | | $\mid$ | | | | | | | | start > 91994 | | | | | | | | | | | | | | start $<=$
104805: translation_(17.47)
| | | | | | | | | | | | | | start > 104805: cell_wall_organization_and_biogenesis_ (15.89)
$|\quad| \quad|\quad|-\left.|\quad|\right|^{-}| |^{-} \mid \quad$ start > 108467:
lipid_metabolic_process_(36.1)
| | ${ }^{-}$| - | ${ }^{-}$| | | start > 113271
| | | | | | | | | | 1 start <= 135384

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=129084$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |

biological_process (28.71)
| | | | | | | | | | | | | start > 122883:
vesicle-mediated_transport_(25.89)

pseudohyphal_growth_(33.74)





```
| | | | | | | | | | | | | start <=
342518: lipid_metabolic_process_ (37.73)
| | | | | | | | - | | | | | | start >
342518: cytokinesis_ (23.29)
```



```
biological_process_ (36.49)
| | | | | | | | | | | | start > 346315:
transcription_(35.34)
| | | | | | | | | | | start > 348523
| | | | | | | | | | start <= 349758:
organelle_organization_and_biogenesis_ (52.84/24.91)
| | | | | | | | | | | | start > 349758:
biological_process_ (20.51)
| | | | | start > 350673
| | | | | start <= 418490
| | | | | | start <= 374871
| | | | | | | | start <= 355044
| | | | | | | | | start <= 350942:
DNA metabolic process (57.91/25.08)
| | | | | | | | | start > 350942
| | | | | | | | | start <= 351717:
biological_process_ (49.61)
| | | | | | | | | | start > 351717
| | | | | | | | | | | start <= 352416:
conjugation_(34.75/14.17)
| | | | | | | | | | | start > 352416:
biological process (18.08)
| | | | | | | | start > 355044
| | | | | | | | start <= 366743
| | | | | | | | | | strand = W
```



```
| | | | | | | | | | | | start <= 357455:
pseudohyphal_growth_(20.49)
| | | | | | | | | | | | | start > 357455:
ribosome_biogenesis_and_assembly_ (24.19)
| | | | | - - | | \ | | | start > 359598:
biological_process_ (28.51)
| lllllllllll
pseudohyphal_growth_(24.52/9.45)
| | | | | | | | | | | | Neigh_GO_aspect = F:
pseudohyphal growth (18.38/6.83)
| | | | | | | | | | | | Neigh_GO_aspect = P:
cytoskeleton_organization_and_biogenesis_(12.93/4.13)
| | | | | | | | | | strand = C:
DNA metabolic process (23.98)
| | | |- | | | | | | start > 366743
| | | | | | | | | start <= 370057:
translation_(27.54)
```


$\mid$ | | | | | | | | $\mid$ start > 394687: biological_process_ (29.17)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>398372$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=408343$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=401042$ |  |

membrane_organization_and_biogenesis_ (22.3) RNA_metabolic_process_(25.95)
| ${ }^{-}$| | ${ }^{-}$| | | | ${ }^{-}$start > 408343:
biological_process_(24.29)
| | | | | | | | | start > 413641
| | | | | | | | | start <= 417304: electron_transport_ (24.32)
| | | | | | | | | | start > 417304:
ribosome_biogenesis_and_assembly_ (25.59)
 DNA_metabolic_process_ (23.71)
| | | | ${ }^{-}$| | | | | |
amino_acid_and_derivative metabolic_process (21.56)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ start > 427737:
DNA_metabolic_process_ (32.67)
| ${ }^{-}| |^{-}| |^{-}|\quad| \quad \mid \quad$ start $>$ 430089: transport (27.21)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>435001$ |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=450873$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid \quad$ start $<=447613$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| \| |  |  |  |  |  |  |  |  |  |  |

biological_process_ (23.26)
| | | | | | | | | | | start > 442360:
cytoskeleton_organization_and_biogenesis_ (18.35)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ | $|\quad| \begin{aligned} & \text { | }\end{aligned}$
$\mid$ | | | | | | | | | | start $<=449870$ :
vesicle-mediated_transport_(19.56)
| | | | $\mid$ | | | | | | | start > 449870:
biological_process_ (28.07)
| | | | | | | | | start > 450873
| | | | | | | | | | start <= 452410:
response_to_stress_ (23.86)
| | $\quad$ - | ${ }^{-} \mid$| | | | 1 start > 452410
| | | | | | | | | | | start $<=$ 456567:
cell_wall_organization_and_biogenesis_ (22.39)
| | | | | | | | | | | start > 456567:
DNA_metabolic_process_(18.62)
| ${ }^{-}$| | ${ }^{-}$| $\left.\right|^{-}$| $\mid$start $>457706$
| | | | | | | | start $<=471379$
$\begin{array}{llllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=467133 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=466336:\end{array}$
biological_process_(45.26)






```
| | | | | | | | | | | | NeighGO term =
signal_transduction_: organelle_organization_and_biogenēsis_(0.0)
| | | | | | | | | | | | NeighGO_term =
ribosome_: organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | | | N̄eighGO_term =
peroxisome_: organelle_organization_and_biogenesis_(0.\overline{0})
| | | | | | - | | | - | - | | NeīghGO_term =
oxidoreductase_activity_: organelle_organization_and_biogenesis_
(0.0)
| | | | | | | | | | | | NeighGO_term =
carbohydrate_metabolic_process_:
organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | | | NeighGO_term =
endoplasmic_reticulum_: organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | | | NeighGO_term =
membrane_: organelle_organization_and_biogenesis_ (0.0)
| | \ | | \ | | | | | | | N
protein_catabolic_process_: organelle_organization_and_\overline{b}iogenesis
(0.0)
| | | | | | | | | | | | NeighGO_term =
translation_regulator_activity_:
organelle_organization_and_biogenesis_ (0.0)
| | | | | | - | - | | | | | | NeighGo_term =
enzyme_regulator_activity_: organelle_organization_and_biogenesis_
(0.0)
| | | | | | | | | | | | NeighGO_term =
membrane_organization_and_biogenesis_:
organelle\overline{e_organization_an\overline{d_biogenesis}_(0.0)}⿻土一⿱幺小
| | | | | | - | - | | | | | | NeighGo_term =
organelle_organization_and_biogenesis_:
organelle_organization_and_biogenesis__ (0.0)
DNA_metabolic_process_: organelle_organization_and_biogenesis_(0.0)
| | | | | | | | | | | | | | - Ne\overline{ighGO_term =}
protein_modification_process_:
organel\overline{le_organization_and_biogenesis_ (0.0)}
| | | | | | | | | | | | NeighGO_term =
transcription_: organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | - | - | Neigh\overline{GO_term =}=0
cell_cortex_: organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | \ | NeighGO_term =
cytoskeleton_: organelle_organization_and_biogenesis_(\overline{0.0)}
| | | | | | | | | | | | NeighGO_term =
protein_kinase_activity_: organelle_organization_and_biogenesis_
(0.0)
| | | | | | | | | | | | NeighGO_term =
cytokinesis_: organelle_organization_and_biogenesis_(0.0)
| | | | | | | | | | | | NeighGO_term =
DNA_binding_: organelle_organization_and_biogenesis_ (0.0)
| | | | | | - | | | \ ` | Neig}hGO_term 
transcription_regulator_activity_:
organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | | | | NeighGO_term =
transferase_activity_: organelle_organization_and_biogenesis_ (0.0)
```

$|||||||\mid$ NeighGO term $=$ Golgi_apparatus_: transport_(1.57/0.52)
$|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGO_term $=$ cell_wall_organization_and_biogenesis_: organelle_organization_and_biogenesis_ (0.0)
$|||||\quad| \quad| \quad| \quad| \quad$ NeighGO_term $=$ chromosome_: organelle_organization_and_biogenesis_(0. $\overline{0}$ )
 phosphoprotein_phosphatase_activity_:
organelle_organization_and_biogenesis_ (0.0)
 response_to_stress_: organelle_organization_and_biogenesis_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGO_term $=$ lyase_activity_: organelle_organization_and_biogenesis_ (0.0)
 cellular_respiration_: organelle_organization_and_biogenesis_ (0.0) $|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGo_term $=$ vesicle-mediated_transport_: transport_ (2.72/1.41)
 peptidase_activity_: organelle_organization_and_biogenesis_ (0.0)
 lipid_metabolic_process_: organelle_organization_and_biogenesis_ (0.0)
$|||||||\mid$ NeighGO_term $=$ plasma_membrane_: organelle_organization_and_biogenesis_(0.0) $|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ | $\mid$ | $\mid$ ighGO_term $=$ cytoskeleton_organization_and_biogenesis_:
organelle_organization_and_biogenesis_ ( $\overline{0} .0$ )
$\left.\left|\left|\left.\right|^{-}\right|\right|\right|^{-}|\quad| \quad \mid \quad$ NeighGo_term $=$ electron_transport_: organelle_organization_and_biogenesis_(0.0)
 conjugation_: organelle_organization_and_biogenesis_ (0.0)
$|||||||\quad| \quad| \quad| \quad$ NeighGO term $=$ vacuole_: organelle_organization_and_biogenesis_(2.37/0.76) $|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGO_term $=$ helicase_activity_: organelle_organization_and_biogenesis_ (6.72/1.39)
$\mid$ | | | | | | | $\mid$ | NeighGO_term = cell_wall_: organelle_organization_and_biogenesis_(0.0) $|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGO term $=$ ligase_activity_: organelle_organization_and_biogenesis_(0.0)
 cellular_bud_: organelle_organization_and_biogenesis_( 0.0 ) $|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ Neigh̄GO term = site_of_polarized_growth_: organelle_organization_and_bīogenesis_ (0.0)
 cell_cycle_: organelle_organization_and_biogenesis_ (0. $\overline{0}$ )
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGO term $=$ isomerase_activity_: organelle_organization_and_biogenesis_ (0.0)
 signal_transducer_activity_: organelle_organization_and_biogenesis_ (0.0)
 amino_acid_and_derivative_metabolic_process_: organelle_organization_and_biogenesis_ (0.0)
 nucleotidyltransferase_activity_:
organelle_organization_and_biogenesis_ (0.0) | | ${ }^{-}$| | $\left.\left.\left.\right|^{-}\right|^{-}|\quad|\right|^{-}|\quad| \quad$ NeighGo_term $=$ microtubule_organizing_center_:
organelle_organization_and_biogenesis_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGo_term $=$ sporulation_: organelle_organization_and_biogenesis_(0.0) $|\quad| \quad-|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGoterm = vitamin_metabolic_process_: organelle_organization_and_bioiogenesis_ (0.0)
 extracellular_region_: organelle_organization_and_biogeñesis_(0.0) $|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad{ }^{-}$NēighGo_term ${ }^{-}=$ nuclear_organization_and_biogenesis_: organelle_organization_and_biogenesis_ (0.0)
 cell_budding_: organelle_organization_and_biogenesis_( $\overline{0} .0)$
| | | | | | | | | NeighGO_term = mitochondrial_envelope_: organelle_organization_and_biogenesis (0.0)
| | | | | | | | | | $\mid$ NeighGO_term =
cytoplasmic_membrane-bound_vesicle_:
organelle_organization_and_biogenesis_ (0.0)
| | | | | | | | | |
biological_process_ (26.63)
$\begin{array}{llllllll}\mid & \mid & \mid & \mid & \mid & \mid & \text { start }>534983 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=555049 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\ \text { start }<=538174 \text { : transport_ }\end{array}$
(57.24/24.42)
$\begin{array}{llllllll}\mid & \mid & \mid & \mid & \mid & \mid & \text { start }>538174 \\ \mid & \mid & \mid & \mid & |\quad| \quad \text { start }<=545269 \text { : }\end{array}$
signal_transduction_(54.05/20.99)
| | | | | | | | | | start > 545269: vesiclemediated_transport_ (42.78/17.4)
$\begin{array}{llllllll}\mid & \mid & \mid & \mid & \mid & \mid & \text { start }>555049 \\ \mid & \mid & \mid & \mid & \mid & \mid & \text { neigh_strand }=\mathrm{W} \text { : }\end{array}$
carbohydrate_metabolic_process_(26.54/10.13)
$|||||||\mid \quad$ neigh_strand $=C$ :
cellular_respiration_(30.58/12.15)
| | $|\quad| \quad \mid \quad$ start > 559815
| | | | | | start $<=614822$
| | | | | | 1 start $<=585588$
| | | | | | | | | start <= 576728
$\begin{array}{llllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=570478 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \\ \text { start }<=560766:\end{array}$
translation_(2.44)
| | | | | | | | | | start > 560766:
biological_process_(34.34)
DNA_metabolic_process_ (30.87)



```
| | | | | | | | | | start > 701897: conjugation
(20.65)
| | | | | | | | | start > 703701: translation_
(28.56)
| | | | | | | start > 706141
| | | | | | | | start <= 707790:
RNA_metabolic_process_(15.61)
```



```
biological_process_ (65.59)
| | | | | | | | | | start > 713657:
RNA_metabolic_process_(7.42/1.3)
| | | | | strrand = C
| | | | | | start <= 706141
| | | | | | | start <= 690323
| | | | | | | start <= 676179
| | | | | | | | | | | start <= 672411 
protein_modification_process_ (17.53)
| | | | | | | | | | start > 668379: translation_
(23.36)
| | | | | | | | | start > 672411: vesicle-
mediated_transport_(48.65/23.83)
| | | | | - | | | start > 676179
| | | | | | | | | start <= 686012
| | | | | | | | | | start <= 678801:
biological_process_ (27.68)
| | | | | | | | | | start > 678801:
response_to_stress_ (32.15)
| | | | | | | | | start > 686012:
protein_modification_process_ (30.27)
| | | | | | | start > 690323
| | | | | | | | start <= 694049:
cytoskeleton_organization_and_biogenesis_(67.32/38.53)
| | | | | | | | start > \overline{694049}
| | | | | | | | | start <= 695597: translation_
(39.07)
| | | | | | | | | start > 695597
| | | | | | | | | start <= 699692: transport_
(27.13)
| | | | | | | | | | start > 699692:
biological_process_ (29.41)
| | | | | | start > 706141
| | | | | | start <= 716122
| | | | | | | start <= 713657: sporulation_
(56.89/25.79)
| | | | | | | | start > 713657:
amino_acid_and_derivative_metabolic_process_ (24.2)
| | | | | | | start > 716122
| | | | | | | | start <= 718329:
protein_modification_process (24.05)
| | | | | | | | start > 718329
| | | | | | | | | start <= 722212:
RNA_metabolic_process_(24.27/3.08)
```




## Decision Tree Generated for Chromosome Fifteen:

## For Cellular Component

start <= 104326
| | start <= 53097
| | | start $<=40747$
| | | | start <= 24293
| | | | | strand = W
| | | | | start <= 6175: cellular_component_ (23.13)
| | | | | | start > 6175: membrane_ (22.91)
$|\quad| \quad|\quad| \quad$ strand $=C \quad$ i $\quad$ start $<=17795$



















## For Molecular Function







| | | | | | | | | | | | | start $<=$ 259490: RNA_binding_ (14.82) | | | | | | | | | | | | | | start > 259490: molecular_function_(9.39)
| | | | | | | | | | | | start > 259924: ligase_activity_ (20.13)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>263476$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| start $<=282165$ |  |  |  |  |  |  |  |
| start $<=271371$ |  |  |  |  |  |  |  |
| $\mid$ | $\mid$ | start $<=265430$ |  |  |  |  |  |

molecular_function_ (9.65)
transferase_activity_ (20.85)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ start $>271371$
| | | | | | | | | | | start <= 277087:
molecular_function_(58.18) | | | | | start > 277087
| | | | | | | | | | | start <= 278057:
RNA_binding_ (23.41)
| | | | | | | | | | | | start > 278057:
molecular_function_(15.87)
| | | | | | | | start > 282165:
transporter activity (27.26)
| | | - | $\quad$ i $\quad$ strand $=\mathrm{C}$
$\begin{array}{lllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=207176 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=198942 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=193542 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=187024 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=181682: \\ \text { hydrolase_activity_ } & (9.9)\end{array}$
| | | | | | | | | | | start > 181682:
RNA_binding_ (20.73)
| | | | | | | | | start > 187024:
hydrolase_activity_ (16.51)
$|||||||\mid$ start > 193542:
molecular_function_(37.33)
| | | | | | | | start > 198942

hydrolase_activity_ (17.59)
| | | | | | | | start > 201879:
transcription_regulator_activity_ (16.98)
| | | | | | | | start > 202518
| | | | | | | | | start <= 204471:
isomerase_activity_ (11.08)
| | | | | | | | |
hydrolase_activity_ (23.69)
| | | | | | | | |
$|\quad| \quad|\quad| \quad|\quad| \quad$ start $<=282165$
| | | | | | | | start $<=231754$
| | | | | | | | | | start < 220766:
molecular_function_(69.52)







```
| | | | | | | | | | start > 515245:
transferase_activity_ (17.27)
| | | | | | start > 525279
| | | | | | start <= 528942:
transcription_regulator_activity_ (14.09)
| | | | | | - | starE > 528942
| | | | | | | start <= 534076
```



```
molecular_function_(14.96)
| | | | | | | | | | start > 530430: RNA_binding_
(11.14)
| | | | | | | | | start > 531509: DNA_binding_
(18.54)
| | | | | | | start > 534076: molecular_function_
(38.58)
| | | start > 539466
| | | | start <= 742911
| | | | | start <= 580251
| | | | | | start <= 559732
| | | | | | start <= 550247
| | | | | | | | strand = W
| | | | | | | | start <= 545030:
peptidase_activity_ (18.9)
| | | | | | | | | start > 545030:
molecular_function_(19.7)
| | | | | | | | strand = C
| | | | | | | | start <= 546858:
nucleotidyltransferase_activity_ (28.23)
| | | | | | | | | start > 546858:
protein_kinase_activity_ (13.45)
| | | | | | | start > 550247
| | | | | | | | start <= 554571
| | | | | | | | strand = W:
oxidoreductase_activity_ (14.23)
| | | | | | | | | strand = C
| | | | | | | | | | start<= 552888:
protein_binding_ (11.98)
| | - | | | | | | | start > 552888:
transcription_regulator_activity_ (17.13)
| | | | | | | | start > 554571
| | | | | | | | | start <= 558643:
peptidase_activity_ (17.93)
| | |- | | | | | | start > 558643:
oxidoreductase_activity_ (17.86)
| | | | | | -start > 559732
| | | | | | start <= 569559
enzyme_regulator_activity_ (14.39)
| | | | \ | | | | strand = C
| | | | | | | | start <= 561171:
hydrolase_activity_ (16.43)
```








| | | | | | | | | | start $<=$ 755329: peptidase_activity_ (10.41)
| | | | | | | | | | | start > 755329:
transferase_activity_ (21.57)

$\mid$ | | | | | | | | | start > 759783:
nucleotidyltransferase_activity_ (17.31)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>761266$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=772602:$ |


| | | | | | | | | $\mid$ start $<=783678$ :
structural_molecule_activity_ (10.61)
| | | | | | | $\mid$ | | | | start > 783678:
molecular_function_(24.86)
$\begin{array}{lllllllll}\mid & \mid & \left.\right|^{-} & \mid & \left.\right|^{-} & \mid & \mid & \text {start }>789858 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \quad \text { start }<=792532:\end{array}$
isomerase_activity_ (14.75)
| | | | | | | | start > 792532:
transferase activity (16.06)
| | | - | $\mid$ | start > 795333
| | | | | | | start $<=848477$
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ start $<=804378$
| | | | | | | | | | start <= 797677:
oxidoreductase_activity_ (22.22)
| | | | | | | | | | | start > 797677:
protein_binding_ (10.8)
| | | | | | | | | | start > 800732
| | | | | | | | | | start <= 802308:
RNA_binding_(12.39)
| | | | | | | | | | start > 802308:
transferase_activity_ (13.06)
$\begin{array}{lllllll}\mid & \mid & \mid & \mid & \mid & \text { start }>804378 \\ \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=841332\end{array}$

transporter_activity $\quad$ (22.91)
$\begin{array}{lllllllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }>807272 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=813984 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\ \text { start }<=\end{array}$
811671: molecular_function_(25.37)
| | | | | | | | | | | | | | start >
811671: peptidase_activity_ (17.57)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ start $>813984$
 824866
| | | | | | | | | | | | | start $<=$ 820453: molecular_function_(8.83)
$\mid$ | | | | | | | | | | | | | | start >
820453: protein_kinase_activity_ (11.07)






## For Biological Process



| | | | | | | | | | | | distance > 2327: RNA_metabolic_process (5.91/2.31)
$\mid$ | | | | | | | | | neigh_num > 1:
transcription_(38.19/20.54)
| | | | | | | | | | start > 53097:
vitamin_metabolic_process_ (16.92)
$\begin{array}{llllllll}\mid & \mid & \mid & \mid & \mid & \mid & \text { start }>56451 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=68754\end{array}$

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=65350$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid \quad$ strand $=W:$ |

biological_process_ $\underset{\mid}{(29.07 / 12.34)} \mid$

translation_(15.8)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ start $>$ 61024:
DNA_metabolic_process_(19.25)
| $\mid$ | $\left.\right|^{-}| |^{-}|\quad| \quad \mid$ start > 65350:
biological_process_(44.96)

cell_cycle_(8.98)
| $\mid$ | | | | | | | | | start > 70325:
cell_wall_organization_and_biogenesis_ (31.51/15.19)
| | | | | | | | | | start > 71300:
biological_process_(11.68)
| | | | | | | | strand = C: transcription_ (10.47)

response_to_stress_ (9.55)
$|\quad| \quad|\quad| \quad|\quad| \quad$ | start > 80348:
carbohydrate_metabolic_process_ (11.62)
$|\quad| \quad|\quad| \quad|\quad| \quad$ start > 83834:
ribosome_biogenesis_and_assembly_ (13.6)
| | $\quad$ | | - |
| | | start <= 91419: transport_ (21.83)
| | | start > 91419
| | | | start <= 94402: translation_ (37.71)
| | | | start > 94402: transport_ (14.26)
$\begin{array}{lllll}1 & \mid & \mid & \mid & \mid \\ \mid & \mid & \text { start }>96361\end{array}$
| | | | | start <= 109176



```
| | | | | | | start > 171070
| | | | | | | | strand =W N start <= 181682:
cell_wall_organization_and_biogenesis_ (19.78)
| | | | | | | | | start > 181682
| | | | | | | | start <= 185438:
generation of precursor metabolites and energy (15.94)
| | | | | | | | | | | | - start > 185438:
cytoskeleton_organization_and_biogenesis_ (18.74)
| | | | | | | | | | | | strand = C | | start < | 187024:
ribosome_biogenesis_and_assembly_ (30.22)
| | | | | - - | | ` start > 187024: transport_
(12.0)
| | start > 193542
| | | | | start <= 212244
| | | | | | start <= 202518
| | | | | | | start <= 201879
| | | | | | | | | | | start <= 198942
| | | | | | | | | | start <= 194970:
DNA metabolic process (16.33)
```



```
sporulation_(17.26)
| | | | | | | | | | | start > 196507:
biological_process_ (12.49)
| | | | | | | | | strand = C: biological_process
(31.51)
| | | | | | | | start > 198942:
DNA_metabolic_process_ (16.22)
| - | | | |- | start > 201879: signal_transduction_
(14.95)
| | | | | | start > 202518
| | | | | | | strand = W: DNA_metabolic_process_
(13.6)
| | | | | | | strand = C
| | | | | | | start <= 204471:
RNA_metabolic_process_ (32.82/15.98)
| | | | | | | | start > 204471
| | | | | | | | | | | | | | start <= 210265 
mediated_transport_ (15.34)
| | \ | | | | | | | start > 205885
| | | | | | | | | | | start <= 207176:
biological_process_(31.5/13.23)
| | | | | | | | | | | start > 207176:
response_to_stress_ (20.56)
| | \ - | - | | | | start > 210265: vesicle-
mediated_transport_ (15.82)
| | | | | start > 212244
| | | | | | start <= 216137:
organelle_organization_and_biogenesis_ (12.86)
```






| | | | | | | | | | start $<=$ 523028: amino_acid_and_derivative_metabolic_process_ (11.03)
| $\left.\left.\left.\left.\right|^{-}\right|^{-}\right|^{-}|\quad|\right|^{-}|\quad|{ }^{-}\left|{ }^{-}\right|$start $>523028$ : cell_wall_organization_and_biogenesis_ (10.78)
 biological_process_ (12.64)
| | | | | | | | | | | | | start > 539466: protein_catabolic_process_(15.19)
| | | | | | | | | | | | start > 545030: biological_process_(18.46)
$\begin{array}{llllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }>546858 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\ \text { ltart }<=554571 \text { : }\end{array}$
response_to_stress_ (10.87)
$|\quad| \quad|\quad|-|\quad| \quad|\quad| \quad$ start $>$ 554571:
pseudohyphal_growth_(11.24)

translation_(19.45)
$\begin{array}{llllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }>571959 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\ \text { start }<=573176:\end{array}$
vesicle-mediated_transport_ (18.55)

translation_(16.61)
| | | | | | | | | | start > 575099:
signal_transduction_ (13.74)

carbohydrate_metabolic_process_(14.46/4.39)



```
| | | | | | | | | | NeighGO term =
biological_process_: conjugation_(0.0)
| | | | | | | | \ | NeighGO_term =
plasma_membrane_: conjugation_(0.0)
| | | | | | | | | NeighGO_term =
transporter_activity_: conjugation_ (0.0)
| | | | | | | | | - | NeighGO_term = transport_:
conjugation_(0.0)
| | | | | | | | | | NeighGO_term =
extracellular_region_: conjugation_ (0.0)
| | | | | | | | | | NeighGO_term =
hydrolase activity : biological process (3.01/\overline{1}.6)
| | | | | | | | | | NeighGO_term =
cell_wall_organization_and_biogenesis_: conjugation_ (0.0)
| | | | | | | | | NeighGO_term = cell_wall_:
conjugation_(0.0)
| | | | | | | | | | NeighGO_term = nucleus_:
signal_transduction_(2.43/1.22)
| | | | | | | | | | NeighGO_term =
transcription_regulator_activity_: conjugation_ (0.0)
| | | | | | | | | | NeighGO_term =
DNA_metabolic_process_: biological_process_(1.0/0.34)
| | | | | |
protein_modification_process_: conjugation_ (0.\overline{0})
| | | | | | | | | | NeighGO_term = chromosome_:
signal_transduction_ (19.78/9.73)
| | | | | | | | | | NeighGO_term = DNA_binding_:
signal_transduction_ (1.57/0.78)
| | | | | - | | | | | NeighGO_term =
RNA_metabolic_process_: conjugation_(0.0)
| | | | | | | | | NeighGO_term =
transcription_
| | | | | | | | | | distance <= 2327:
biological_process_(3.06/0.55)
| | | | | | | | | | distance > 2327:
conjugation_(2.28/0.93) | | | | | | | | | NeighGO_term = nucleolus_:
conjugation_(1.32/0.27)
| | | | | | | | | | NeighGO_term =
ribosome_biogenesis_and_assembly_: conjugation_- (0.0)
| | | | | | | | | NeighGO_term = cytoplasm_:
conjugation_(0.0)
| | | | | | | | | | NeighGO_term =
isomerase_activity_: conjugation_ (0.0)
| | | | | | | | \ | NeighGO_term =
vitamin_metabolic_process_: conjugation_ (0.0)
| | | | | | | | | | NeighGO_term = RNA_binding_:
conjugation_(0.0)
| | | | | | | | | | NeighGO_term = ribosome_:
conjugation (0.0)
| | | | | | | | | | NeighGOterm =
translation_regulator_activity_: conjugation_(\overline{0.0)}
| | | | | | | | | | NeighGO_term = translation_:
conjugation_(0.0)
```


| | | | | | | | $\mid$ NeighGO_term = peptidase_activity_: conjugation_(0.0)
| | | | | | | | $\mid$ NeighGO_term = protein_catabolic_process_: conjugation_ (0.0) | | | | | | | | | | NeighGO_term = vesiclemediated_transport_: conjugation_(0.0)
| | $\mid$ | | | | | $\mid$ | NeighGO_term = lipid_metabolic_process_: conjugation_ (0.0)
| | | | | | | | NeighGO_term = cellular_respiration_: conjugation_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad$ | $\mid$ NeighGo_term $=$ cellular_bud_: conjugation_(0.0)
$|\quad| \quad|\quad| \quad \mid \quad$ | $\mid \quad$ NeighGO term $=$ site_of_polarized_growth_: conjugation_ (0.0)
| | | | | | | | NeighGO_term = enzyme_regulator_activity_: conjugation_(0.0)
| | | | | | | | | | NeighGO_term = mitochondrial_envelope_: conjugation_(0.0)
| | | | | | $\mid$ | | $\mid$ NeighGO_term = pseudohyphal_growth_: conjugation_(0.0)
| | | | | | | | | NeighGO_term = helicase_activity_: conjugation_(0.0)
$|\quad| \quad|\quad| \quad \mid \quad$ | $\mid \quad$ NeighGO_term $=$ cytoskeleton_: conjugation_(0.0)
| | | | | | | | | NeighGO_term = electron_transport_: conjugation_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ NeighGO term $=$ membrane_organization_and_biogenesis_: conjugation_ (0.0)
$|\quad| \quad|\quad| \quad|\quad|-\quad|\quad| \quad$ NeighGo_term = cellular_homeostasis_: conjugation_ (0.0)
| | | | | | | | | NeighGO_term = nucleotidyltransferase_activity_: conjugation_(1.96/0.48)
| | | | | | | | | NeighGO_term = meiosis_: conjugation (0.0)
| | | - | | | | | | NeighGO_term =
amino_acid_and_derivative_metabolic_process_: conjugation_ (0.0)
| | | | | | | | | NeighGO_term =
phosphoprotein_phosphatase_activity_: conjugatiōn_(0.0)
| | | | | | | | | | NeighGo_tērm = sporulation_: conjugation_(0.0)
| | | | | | | | | NeighGO_term =
motor_activity_: conjugation_(0.0)

conjugation_(0.0)
| | | - | | | | | | NeighGO_term = peroxisome_:
conjugation_(0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad$ NeighGO_term $=$
nuclear_organization_and_biogenesis_: conjugation_ (0.0)
$|\quad| \quad|\quad| \quad|\quad| \quad$ - $|\quad| \begin{array}{ll}\text { start }>74528 \overline{2}\end{array}$
| | | | | | | | | start <= 749302
$\mid$ | | | | | | | | | neigh_strand = C:
DNA_metabolic_process_(11.12/3.96)
$\mid$ | | | | | | | | | | neigh_strand $=\mathrm{W}$ :
cell_cycle_(12.05/3.45)











```
| | | | | | | | | | | | start > 712867:
biological_process_ (15.28)
| llllllllllllll
meiosis (10.54)
| | | | | | | | | | | | start > 720066:
ribosome_biogenesis_and_assembly_(28.03/11.06)
| | | | | | - | | | start > 722912
| | | | | | | | | | start <= 745282
```



```
| | | | | | | | | | <= 727513
biological_process_ (11.61)
| | | | | | | | | | | | | start > 727513:
transcription_ (13.15)
| | | | | | | | | | | | start > 733926:
biological_process_ (38.53/17.7)
| | | | | | | | | | | | start > 738321
| | | | | | | | | | | | start<= 742911:
organelle_organization_and_biogenesis_ (25.2)
| | | | | | | | | | | | start > 742911:
transcription (17.32)
| | | |- | | | | | | start > 745282:
biological_process_ (34.04)
| | | | | | | | start > 747283
```



```
(37.84/13.38)
| | | | | | | | | | start > 749302:
protein_modification_process_ (14.19)
| | | | | \ | \ | start > 755329
| | | | | | start <= 758331:
cellular_respiration (11.41)
| | \ | | \ | | | | | start > 758331:
ribosome_biogenesis_and_assembly_ (31.99/15.02)
| | | | | | | | | | start > 759783:
transcription (16.21)
| | | | | | start > 761266
| | | | | | start <= 800732
| | | | | | | start <= 789858
| | | | | | | | | start <= 772602:
biological_process_(39.03/11.38)
| | | | | | | | | start > 772602
| | | | | | | | | | start <= 783678:
translation_(9.24)
| | | | | | | | | | start > 783678:
cell_wall_organization_and_biogenesis_ (35.88/15.58)
| | | | | | | | start > 789858
| | | | | | | | start <= 795333
| | | | | | | | | start <= 792532:
ribosome_biogenesis_and_assembly_ (25.38)
| | | | | | | | | | start > 792532:
lipid_metabolic_process_ (21.48)
```


| | | | | | | | | | | start > 851838: biological_process_ (15.54)

(33.95/8.51)
| | | | | | | | start > 894089:
RNA_metabolic_process_(11.32)

(21.36/8.47)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>901191$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=907552$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid \quad$ start $<=904757:$ |

sporulation_(11.42)
| | | | | | | | | | start > 904757:
transport (13.54)
| | | | | | | | | start > 907552:
biological_process_ (15.51)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>909340$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=944593$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ start $<=925718$ |  |  |  |
| $\mid$ | start $<=916027:$ |  |  |  |  |  |

protein_modification_process_(15.32)
| | | | | | | | | | start > 916027:
DNA_metabolic_process_(17.02)
| ${ }^{-}$| | ${ }^{-}$| | | | start > 921059 | | | | | | | | $\mid$ start <= 922902:
amino_acid_and_derivative_metabolic_process_ (19.6)
| | | | | | | | | start > 922902:
response_to_stress_ (11.86)




## Decision Tree Generated for Chromosome Sixteen:

## For Cellular Component





```
| | | | | | | | | NeighGO term = nucleolus :
nucleus (0.0)
| | | | | | | | | | NeighGO_term =
vitamin_metabolic_process_: nucleus_(0.0)
| | | | | | | | | NeighGO_term = RNA_binding_:
nucleus (0.0)
| | | | | | | | | | NeighGO_term =
DNA_metabolic_process_: nucleus_ (0.0)
| | | | | | | | | NeighGO_term =
isomerase_activity_: nucleus_ (0.0)
| | | | | | | | | | NeighGO_term =
lipid_metabolic_process_: nucleus_ (0.0)
| | | | | | | | | | NeighGO_term =
plasma_membrane_: nucleus_(0.0)
| | | | | | | | | | NeighGO_term =
protein_kinase_activity_: nucleus_ (0.0)
| | | | | | | | | NeighGO term = DNA binding :
nucleus (0.0)
| | | | | | | | | | NeighGO_term =
transcription_regulator_activity_: nucleus_ (0.0)
| | | | | | | | | | NeighGO_term = meiosis_:
nucleus_ (0.0)
| | | | | | | | | | NeighGO term = ribosome :
nucleus_ (0.0)
| | | | | | | | | NeighGO_term = translation_:
nucleus_ (0.0)
| | | | | | | | | | NeighGO_term =
nuclear_organization_and_biogenesis_: nucleus_ (0.0)
| | | | | \ \ | | - | Neigh\overline{GO_term =}
transcription_: nucleus_ (0.0)
| | | | | | | | | | NeighGO_term =
nucleotidyltransferase_activity_: nucleus_ (0.0)
| | | | | | | | | | NeighGO_term =
ligase_activity_: nucleus_ (0.0)
| | | | | | | | | | NeighGO_term = peroxisome_:
nucleus_(0.0)
| | | | | | | | | NeighGO_term =
mitochondrial_envelope_: nucleus_(0.0)
| | | | | | | | | \ | NeighGO_term =
extracellular_region_: nucleus_ (0.0)
| | | | | | | | | | NeighGO_term =
anatomical_structure_morphogenesis_: nucleus_(\overline{0}.0)
| | | | | | | | | | NeighGO_term =
membrane_fraction_: nucleus_ (0.0)
| | | | 
pseudohyphal_growth_: nucleus_ (0.0)
| | | \ | | | | | | NeighGO_term =
amino_acid_and_derivative_metabolic_process_: nucleus_ (0.0)
| | | | | | | | | | NeighGO_term =
protein_catabolic_process_: nucleus_ (0.0)
| | | | | | | | | - | NeighGO_term =
oxidoreductase_activity_: nucleus_ (0.0)
| | | | | | | | | | NeighGO_term =
Golgi_apparatus_: nucleus_(0.0)
```

 (16.16)




















|  |  |  |
| :---: | :---: | :---: |
|  |  |  |
| \| | | \| | | | NeighGo_term = transferase_activity_: |
| ribosome_ (0.0) |  |  |
| 1 \| |  | NeighGo_term = |
| protein_modification_process_: $\overline{\text { ribosome_ }}$ (0.0) |  |  |
| \| | - | \| | | | NeighGo_term = cellular_homeostasis_: |
| ribosome_ (0.0) |  |  |
| 1 \| 1 | \| | | | NeighGO_term = nucleolus_: ribosome_ (0.0) |
| 1 \| | | \| | | | NeighGo_term = vitamin_metabolic_process_: |
| ribosome_ (0.0) |  |  |
| 1 \| | | \| | | NeighGo_term = RNA_binding_: ribosome_ (0.0) |
| 1 \| | 1 \| | | NeighGO_term = DNA_metabolic_process_: |
| ribosome_ (0.0) |  |  |
| \| | | \| | | | NeighGO_term = isomerase_activity_: |
| ribosome | (0.0) |  |
| \| | | 1 \| | NeighGO_term = lipid_metabolic_process_: |
| ribosome | (0.0) |  |
| 1 \| | 1 \| | NeighGo_term = plasma_membrane_: ribosome |
| (3.21/0.8) |  |  |
| \| | 1 \| | NeighGO_term = protein_kinase_activity_: |
| ribosome | (0.0) |  |
| 1 \| | | 1 \| | NeighGo_term = DNA_binding_: ribosome_ (0.0) |
| 1 \| | | 1 \| | | NeighGO_term = |
| transcription_regulator_activity_: ribosome_ (0.0) |  |  |
| \| | | |  | NeighGo_term = meiosis_: ribosome_ (0.0) |
| 1 I | \| | NeighGO_term = ribosome_: ribosome_ (0.0) |
| 1 I 1 | 1 \| | NeighGO_term = translation_: ribosome_ (0.0) |
| 1 \| | | 1 \| | | NeighGo_term |
| nuclear_organization_and_biogenesis_: ribosome_ (0.0) |  |  |
|  | \| | | | NeighGo_term = transcription_: ribosome_ |
| (0.0) |  |  |
| 1 \| | | \| | | | NeighGO_term = |
| nucleotidyltransferase_activity_: ribosome_ (0.0) |  |  |
|  | \| | | | NeighGo_term = ligase_activity_: ribosome_ |
| (1.57/0.35) |  |  |
| 1 \| | | 1 \| | NeighGO_term = peroxisome_: ribosome_ (0.0) |
| 1 \| | 1 \| | | NeighGo_term = mitochondrial_envelope_: |
| ribosome_(8.65/1.09) |  |  |
| \| | |  | NeighGo_term = extracellular_region_: |
| ribosome | (0.0) |  |
| \| | \| | | | NeighGO_term = |
| anatomical_structure_morphogenesis_: ribosome_ (0.0) |  |  |
|  | \| | | | NeighGo_term = membrane_fraction_: ribosome_ |
| (10.01/0.47) |  |  |
| \| | | 1 \| | | NeighGO_term = pseudohyphal_growth_: |
| ribosome_ (0.0) |  |  |
|  |  | NeighGO_term = |
| amino_acid_and_derivative_metabolic_process_: cytoplasm_ (2.49/0.35) |  |  |
| \| $\left.\left.\left.\right\|^{-}\right\|^{-}\right\|^{-}$\| | |  | NeighGo_term = protein_catabolic_process_: |
| ribosome_(1.08/0.46) |  |  |
| \| | | | 1 \| | NeighGO_term = oxidoreductase_activity_: |
| ribosome_ | (0.0) |  |







## For Molecular Function

| start <= 634120
| | stop <= 518732
| | | start <= 138697
| | | | start $<=63279$

```
            start <= 63279
            | start <= 11887
```

            | strand \(=C\)
                        | start \(<=8427\)
    


```
oxidoreductase_activity_ (15.05)
| | | | | | | | | start > 73881
| | | | | | | | | start <= 74719:
molecular_function_(7.4)
| | | | | | | | | | start > 74719: RNA_binding_
(9.21/1.32)
| | | | | | | distance > 1665
| | | | | | | | distance <= 1860:
motor activity (10.14/1.24)
| | | | | | | | | distance > 1860:
enzyme_regulator_activity_ (37.19/3.42)
| | | | | | | strand = W
| | | | | | | start <= 67725:
structural molecule activity (20.24)
| | | | | | | | start > 67725:
transcription_regulator_activity_ (11.01)
| | | | | start > 79353: DNA_binding_(39.08/17.84)
| | | | | start > 82356
| | | | | start <= 107275
| | | | | | strand = C
| | | | | | | start <= 100496
| | | | | | | | | | | | | | | | | | start <= 88517 
molecular_function_(24.06)
| | | | | | | | | | start > 85586:
transporter_activity_ (18.84)
| | | | | | | | | start > 88517
```



```
protein_binding_ (17.95)
| | | | | | | | | | | start > 95109:
molecular_function_(18.56)
| | | | | | | | | | start > 96258:
protein_binding_ (18.79)
| | - | - | | | | start > 100496
| | | | | | | | | start <= 103232:
protein_kinase_activity_ (9.25)
| | | | | | | | start > 103232:
transporter_activity_ (35.14/14.59)
| | | | | | | strand = W
| | | | | | | start <= 106172
| | | | | | | | start <= 99484:
molecular_function_(38.2)
| | | | | | | | | start > 99484
| | | | | | | | | | start <= 103232
```

```
| | | | | | | | | | start <= 100496:
translation_regulator_activity_ (11.2)
| | | | | | | | | | | start > 100496:
helicase_activity_ (21.98)
| | | | | | | | | | start > 103232:
molecular function (15.24)
| | | | | | | | start > 106172: protein_binding_
(19.29)
\begin{tabular}{lllll|l}
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & start \(>107275\) \\
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & start \(<=130161\) \\
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & start \(<=108652:\)
\end{tabular}
transferase_activity_(41.39/20.65)
|
transferase_activity_ (20.81)
| | | | | | | | start <= 117067:
molecular_function_(29.65)
| | | | | | | | | | | start > 117067:
hydrolase_activity_ (56.58)
| | | | | | | | | start > 121767:
molecular_function_(67.4)
| | | | | | | start > 130161
| | | | | | | start <= 135789
| | | | | | | | | start <= 133042:
transporter_activity_ (18.3)
| | | | | | | | | start > 133042:
structural_molecule_activity_ (21.08)
| | | | | | | | start > 135789
| | | | | | | | | start <= 136749:
enzyme_regulator_activity_ (16.69)
| | | | | | | | | start > 136749:
hydrolase_activity_(20.77)
| | | start > 138697
| | | | start <= 289033
| | | | | start <= 230837
| | | | | | start <= 169337
| | | | | | start <= 159908
| | | | | | | | strand = C
| | | | | | | | start <= 156211
```



```
molecular_function_(9.18)
| | | | | | | | | | | | start > 143820:
transferase_activity_ (16.55)
| | | | | | | | | | start > 150613
| | | | | | | | | | | start<= 153494:
isomerase_activity_ (13.75)
| | |- | | | | | | | | start > 153494:
molecular_function_(17.64)
| | | | | | | | | start > 156211:
protein_kinase_activity_ (21.48)
| | | | | | | | strand = W
```



```
| | | | | | | | | | | start <= 175042:
molecular_function_(12.63)
| | | | | | | | | | | | start > 175042:
protein_binding_ (18.07)
| | | | | | | | | | | | start > 176222:
molecular function (31.05)
| | | | | | | | | | | start > 184678
| | | | | | | | | | start <= 191405:
transferase_activity_ (27.51)
```



```
| | | | | | | | | | | start <= 195425:
protein_binding_ (18.69)
| | | | | | | | | | | | | start > 195425
| | | | | | | | | | | | | | start <=
199094: structural_molecule_activity_ (13.08)
| | | | | | | | | | | | | | start >
199094: protein_binding_ (7.97)
| | | | | | - | | | | | | start > 203420:
molecular_function_(13.63)
| | | | | | | | | start > 205247
| | | | | | | | | | start <= 208156:
phosphoprotein_phosphatase_activity_ (11.02)
| | | | | | | | | - | start > 208156: RNA_binding_
(17.83)
| | | | | | | start > 218362
| | | | | | | | strand = C
| | | | | | | | | | | | | | | | | | | | | start <= 225740 
protein_binding_ (20.9)
| | | | | | | | | | start > 223142:
transferase_activity_ (14.28) , start > 225740
| | | | | | | | start <= 228313:
oxidoreductase_activity_ (26.4)
| | | | | | | | | | start > 228313:
protein_binding_ (27.25/11.63)
| | | | | | | | strand = W
| | | | | | | start <= 223142
| | | | | | | | | | start <= 218629:
transferase_activity_ (11.78)
| | | | | | | | | | start > 218629:
structural_molecule_activity_ (19.9)
| | | | | | | | | start > 223142:
molecular function (14.87)
| | | | | start > 230837
| | | | | | start <= 254309
| | | | | | | | | | start <= 237106 
(24.08)
| | | | | | | | start > 231216:
nucleotidyltransferase_activity_ (20.93)
| | | | | | | start > 237106
```



```
| | | | | | | | | strand = C:
protein_kinase_activity_ (25.03)
| | | | | | | | | strand = W:
structural_molecule_activity_ (7.57)
| | | | start > 289033
| | | | start <= 335485
| | | | | | start <= 310209 | | start <= 301227
| | | | | | | start <= 297552
| | | | | | | | | start <= 291050:
transcription_regulator_activity_ (21.51)
| | | | | | - | | | start > 291050
```



```
protein_binding_ (14.57)
| | | | | | | | | | | start > 292426:
enzyme_regulator_activity_ (13.67)
| | | | \ | | | | | strand = W: protein_binding_
(17.41)
| | | | | | | | start > 297552
| | | | | | | | | start <= 299502:
transporter_activity_ (18.52)
| | | | | | | | start > 299502:
transcription_regulator_activity_ (20.09)
| | | | | | - | start > 301227
| | | | | | | strand = C: DNA_binding_ (28.96)
```



```
| | | | | | | | | | start <= 301715:
molecular_function_(25.18)
| | | | | | | | | | start > 301715
| | | | | | | | | | | start <= 303120:
RNA_binding_(17.21)
| | | | | | | | | | | start > 303120:
molecular_function_(17.65)
| | | | | | | | | start > 304386
| | | | | | | | | | start <= 305297:
transcription_regulator_activity_ (11.35)
| | | | | | - | | | | start > 305297: RNA_binding_
(20.9)
\begin{tabular}{llllllll}
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & start \(>310209\) \\
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & start \(<=326627\) \\
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & strand \(=C\) \\
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & start \(<=320766\) \\
\(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & \(\mid\) & start \(<=318944:\) \\
hydrolase activity & \((17.94)\)
\end{tabular}
| | | | | | | | | | start > 318944:
transcription_regulator_activity_ (14.39)
| | | | | | | | | start > 320766
| | | | | | | | | | start <= 324286:
molecular_function_(30.61)
| | | | | | | | | | start > 324286:
helicase_activity_ (14.91)
| | \ | | | | | strand = W
```














```
| | | | | | | | | | | | | | | | | | start <= 895957:
hydrolase_activity_ (17.81)
| | | | | | | | | start > 895957:
enzyme_regulator_activity_ (16.76)
| | | | | | strand = W
| | | | | | start <= 888970
| | | | | | | start <= 885792
| | | | | | | | | start <= 880365:
molecular_function_(22.21)
```



```
protein_binding_ (13.39)
| | | | | | | | | | start > 883824:
molecular_function_(15.08)
| | | | | | | | start > 885792:
nucleotidyltransferase activity (17.14)
| | | | | | - | starrt > 888970
| | | | | | | start <= 895957
| | | | | | | | | start <= 892328: RNA_binding_
(14.96)
| | | | | | | | | start > 892328:
molecular_function_(17.15)
| | | | | | | | start > 895957: RNA_binding_(23.77)
| | start > 900190
| | | start <= 919037
| | | | start <= 902040: transferase_activity_ (60.17/17.59)
| | | | | | start > 902040
| | | | | | start <= 907214: protein_binding_ (16.81)
| | | | | | start > 907214:
transcription_regulator_activity_ (19.14)
| | | | | start > 911018
| | | | | | strand = C
| | | | | | | start <= 912660: protein_binding_
(18.97)
| | | | | | | start > 912660:
nucleotidyltransferase_activity_ (22.44)
| | | | | | strand = W
| | | | | | | start <= 911253:
nucleotidyltransferase_activity_ (22.59)
| | | | | | start > 911253: molecular_function_
(13.75)
| | | start > 919037
| | | | strand = C
| | | | | start <= 931372
| | | | | | start <= 923375: transferase_activity_
(18.24)
| | | | | | start > 923375: transporter_activity_
(12.76)
| | | | | start > 931372
| | | | | | start <= 939028:
transcription_regulator_activity_ (20.18)
```

```
| | | | | | start > 939028: oxidoreductase_activity_
(12.57)
| | | | strand = W
| | | | | start <= 919377: oxidoreductase_activity_
(20.91)
| | | | | start > 919377
| | | | start <= 939918
| | | | | | start <= 931372
| | | | | | | | start <= 923375:
transporter_activity_ (9.77)
| | | | | \ | | start > 923375: molecular_function_
(10.22)
| | | | | | | start > 931372: transporter_activity_
(16.7)
| | | | | | start > 939918
| | | | | | start <= 943876: molecular_function_
(15.82)
| | | | | | start > 943876: helicase_activity_
(6.54)
```


## For Biological Process

start $<=96258$
$\mid$
$\mid$



| | | | | | | | | start > 164275:
protein_modification_process_ (16.9)

| $\mid$ | $\mid$ | - | $\mid$ | start $>\overline{1} 66255$ |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | strand $=C$ |  |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=171483$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid \quad$ start $<=169337$ : response_to_stress_ |
| $(14.87)$ |  |  |  |  |  |  | (14.87)

| | | | | | | | start > 169337:
generation_of_precursor_metabolites_and_energy_ (11.06)
$\begin{array}{llllllll}\mid & \mid & \mid & \mid & \mid & \mid & \text { start }>171483 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=183055 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\ \text { start }<=176222:\end{array}$
biological_process_ (16.98)
$\begin{array}{lllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }>176222 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \quad \text { Neigh_GO_aspect }=C:\end{array}$ nuclear_organization_and_biogenesis_ (11.19/4.08)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ । $\mid$ Neigh_GO_aspect $=\mathrm{F}$ : nuclear_organization_and_biogenesis_ (4.71/2.04)
$|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ $\mid \quad$ Neigh_GO_aspect $=P$ :
conjugation_(11.28/1.48)
$\begin{array}{lllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \text { start }>183055 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=184678:\end{array}$
biological_process_ (15.04)
| | | $\left.{ }^{-1}\right|^{-}|\quad| \quad|\quad| \quad$ start $>184678$
| | | | | | | | | start <= 187724:
transcription_ (7.33)
| | | | | | | | | | start > 187724:
biological_process_ (12.43)
| | | ${ }^{-}$| $\left.\right|^{-}$strand $=W$
| | | | | | start $<=184678$
| | | | | | | start <= 179276
| | | | | | | | $\operatorname{start}<=173826$
$\left\lvert\, \begin{array}{llllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=173151 \\ \mid & \mid & \mid & \mid & \mid & \mid & \text { start }<=171483:\end{array}\right.$
meiosis_(16.62)

translation_(9.02)
| | | | | | | | | start > 173151:
ribosome_biogenesis_and_assembly_(26.34/12.46)
| | $\quad|\quad| \quad-\quad$ - $|\quad| \quad$ | start > 173826
| | | | | | | | | | start <= 176222
| | | | | | | | | | | start <= 175042:
response_to_stress_ (11.23)
$|\quad| \quad|\quad|-|\quad| \quad|\quad| \quad$ start > 175042: vesicle-
mediated_transport_ (10.77)
| | $\mid$ | | | | | | | start > 176222: meiosis_
(25.13)
| | | | | | | | start > 179276
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad$ neigh_strand $=W$ : DNA_metabolic_process_(11.61/3.75)
| | | | | | | | neigh_strand $=C$ | | | | | | | | | distance <= 4146:
DNA_metabolic_process_(6.78/2.87)




```
| | | | | | | | | | | | | | start <=
326627
| | | | | | | | | | | | | | start <=
324286: biological_process_(16.98) | | | | | | | >
324286: translation (15.06)
| | | | | - | | | | | | | | | start >
326627
| | | | | | | | | | | | | | | start <=
329605: lipid_metabolic_process_ (19.75)
| | | | | | | | | | | | | | | start >
329605
| | | | | | | | | | | | | | | | |
start <= 335485: pseudohyphal_growth_ (15.56)
| | | | | | | | | | | | | | | |
start > 335485: biological_process_ (11.31)
| | | | | | | | | | | | | start > 337142
| | | | | | | | | | | | | | start<=
338619: organelle_organization_and_biogenesis_ (15.86)
| | | | | | | | | | | | | | start >
338619
| | | | | | | | | | | | | | | start <=
339943: translation (19.28)
| | | | | - | | | | | | | | | | start >
339943: lipid_metabolic_process_ (10.34)
| | | | | | | | | | | | start > 344738
| | | | | | | | | | | | start <= 364065
352272
| | | | | | | | | | | | | | | start <=
349116: biological_process_ (18.33)
349116: DNA metabolic process (15.5)
| | | | | | | |- | | | | | | start >
352272: biological_process_ (43.09)
| | | | | | | | | | | | | start > 364065:
organelle_organization_and_biogenesis_ (20.67)
| | | | | | | | | | | start > 364949
| | | | | | | | | | | | start <= 369438
| | | start <= 366924
protein_modification_process_ (14.31)
| | | | | | | | | | | | | start > 366924:
lipid_metabolic_process_ (17.7)
| | | | | | | | | | | | start > 369438
| | | | | | | | | | | | | start <= 373790:
transport (13.96)
| | | | | | | | | | | | | start > 373790:
translation_(22.72)
| | | | | | | | | | start > 378392
| | | | | | | | | start <= 408741
| | start <= 383452:
cell_wall_organization_and_biogenesis_ (42.39/18.75)
| | | | | | | | | | | | start > 383452
```

| | | | | | | | | | | | start <= 398102: RNA_metabolic_process_ (40.31)

453054: biological_process_ (18.73)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>453732$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=461963$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid \quad$ start $<=455756:$ |

protein_modification_process_(46.02/28.47)
| | - | | $|\quad| \quad|\quad| \quad|\quad| \quad$ start > 455756
| | | | | | | | | | | start $<=459960$ :
cell_wall_organization_and_biogenesis_ (17.17)

cytokinesis_(24.98/8.35)
$\begin{array}{llllllllll}\mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }>461963 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\ \text { start }<=464398:\end{array}$
DNA_metabolic_process_ (15.31)
$\begin{array}{lllllllllll}-- & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \text { start }>464398 \\ \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid & \mid \\ \text { start }<=469936: ~\end{array}$
RNA_metabolic_process_ (13.93)
| ${ }^{-}| |^{-}| |^{-}|\quad| \quad|\quad| \quad \mid \quad$ start $>469936$ :
meiosis_(14.06)
| | - | | | | start > 474704
| | | | | | | start $<=560286$
| | | | | | | | start <= 539382
| | | | | | | | | 1 start <= 514176
| | | | | | | | | $\mid$ start $<=491362$
| | | | | | | | | | | start $<=480179$
| | | | | | | | | | | | start <= 475738:
DNA_metabolic_process_ (20.34)
| | | | | | | | | | | | | start > 475738:
translation_(12.44)
| | | | | | | | | | | start > 480179
$\mid$ | | | | | | | | | | start $<=$ 488205:
biological_process_ (35.26)
| | | | | | | | | | | | | start > 488205:
vesicle-mediated_transport_ (18.84)
| | | | $\quad|\quad| \quad|\quad| \quad|\quad| \quad$ start $>491362$









```
| | | | | | | | strand = C:
amino_acid_and_derivative_metabolic_process_ (17.84)
```



```
(24.67/11.58)
| | | | | | | | | start > 685895:
biological_process_ (19.35)
| | | | | - start > 688169
```



```
| | | | | | | start <= 692415:
protein_modification_process_ (15.01)
| | | | | \ | \ start > 692415
| | | | | | | | | neigh_num <= 1:
vitamin_metabolic_process_(11.18/2.63)
| | | | | | | | | neigh_num > 1:
carbohydrate_metabolic_process_(25.32/10.99)
| | | | | | \ | stārt > 694835: response_to_stress_
(28.96/10.86)
| |llllllll
(12.95)
| | start > 696816
| | | start <= 833685
| | | | start <= 814054
| | | | | start <= 760023
| | | | | strand = C
| | | | | | start <= 732347
| | | | | | | start <= 719240
```



```
| | | | | | | | | | start<= 700592:
biological_process_ (16.3)
```



```
translation_(16.14)
| | | | | | | | | | | | start > 703968:
RNA_metabolic_process_ (14.36)
```



```
| | | | | | | | | | | | start <= 710099:
biological_process_ (12.0)
| | | | | | | | | | | | start > 710099:
transport_ (7.77)
| | | | | | | | | | | start > 713273:
biological_process_ (18.45)
| | | | | | | | | start > 718466:
ribosome_biogenesis_and_assembly_(110.23/52.49)
| | | | | - - | | start > 719240
| | | | | | | | start <= 730490
| | | | | | | | | start <= 724715
| | | | | | start <= 720635:
```




| | | | | | | | | | NeighGO_term = sporulation_: vesicle-mediated transport_ (0.0)
| | | | $\rceil$ | | ${ }^{-}$| | NeighGO_term =
phosphoprotein_phosphatase_activity_: vesicle-mediated_transport_ (0.0)
| | | | | | | | | | NeighGO_term = electron_transport_: vesicle-mediated_transport- (0.0) | | | | | | | | | | NeighGO_term = membrane_organization_and_biogenesis_: vesicle-mediated_transport_ (0.0)
| | | | | | | | | | NeighGO_term = cell_cortex_:
vesicle-mediated_transport_ (0.0)
| | | | | | | ${ }^{-}$| | NeighGo_term =
translation_regulator_activity_: vesicle-mediated_transport_ (0.0)
| | | | | | | | | | NeighGO_term =
cell_budding_: vesicle-mediated_transport_ (0.0)
| $\overline{1} \quad|\quad| \quad|\quad| \quad \mid \quad$ start > $7 \overline{4} 0059$
| | | | | | | | | start <= 744282:
RNA_metabolic_process_ (17.6)
| | | | | | | | | start > 744282:
ribosome_biogenesis_and_assembly_ (34.32)

| | | | | | | | start $<=700592$
| | | | | | | | | start <= 698867: transport_
(8.65)
| | | | | | | | | start > 698867: translation_
(11.57)
| | | | | | | | start > 700592:
organelle_organization_and_biogenesis_ (16.7)


| | | | | | | | | organelle_organization_and_biogenesis_ (9.85)
 cytoskeleton_organization_and_biogenesis_ (10.09)
$|\quad| \quad|\quad| \quad|\quad| \quad|\quad| \quad \mid \quad$ strand $=W$
| | | | | | | | | | start <= 792683: RNA_metabolic_process (10.75)
| | | | | | | | | | | start > 792683: translation_(15.41)

| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>794961$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $<=797553$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ |
| start $<=797206:$ |  |  |  |  |  |  |  |  |  |

RNA_metabolic_process_(14.41)
| | | | | | | | | | start > 797206:
transport_(28.52/11.54)
| | | | | | | | | | start > 797553:
RNA_metabolic_process_(17.78)

| $\mid$ | - | $\mid$ | $\mid$ | $\mid$ | $\mid$ | start $>798047$ |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | strand $=C$ |
| $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid$ | $\mid \quad$ start $<=812449:$ |

biological_process_(10.69)
| | | | | | | | | start > 812449: transport
(14.95)
| | | | | | | | strand = W
| | | | | | | | start <= 802355
| | | | | | | | | | start <= 799230:
DNA_metabolic_process_ (11.33)
| | | | | | | | | | start > 799230:
ribosome_biogenesis_and_assembly_ (13.02)
| | | | | | | | | start > 802355: translation_
(12.01)
| | | | start > 814054
| | | start $<=821419$
| | | | | start $<=$ 814387: electron_transport_
(40.87/13.53)
| | | | | | start > 814387
$\mid$ | | | | $\mid$ start $<=817919$ : meiosis_ (18.02)
| | | | | | | start > 817919
| | | | | | | distance <= 1187: translation_
(15.38/2.4)
| | | | | | | distance > 1187:
ribosome_biogenesis_and_assembly_ (15.1/1.07)
| | $\quad$ | | -star̄t > 8214 19
| | | | | | start $<=822616$
| | | | | | start <= 821728: biological_process_
(14.3)
| | | | | | | start > 821728:
amino_acid_and_derivative_metabolic_process_ (34.32)
| | | | | | start > 822616
| | | | | | start <= 829914: biological_process_
(77.96)
| | | | | | | start > 829914
| | | | | | | start <= 831671:
protein_catabolic_process_ (12.28)


```
| | | | | | | | start > 843258
| | | | | | | | start <= 860411
| | | | | | | | | | start <= 856898:
translation_(14.1)
| | | | | | | | | | start > 856898
| | | | | | | | | | | start <= 857579:
cell_wall_organization_and_biogenesis_ (9.86)
| | | | | | | | | | | start > 857579:
translation_(7.3)
| | | | | | | | start > 860411:
generation_of_precursor_metabolites_and_energy_ (14.46)
| | | - | | | - | start > - 861\overline{3}02
| | | | | | | | start <= 880365
```



```
DNA_metabolic_process_ (11.12)
| | | | | | | | | | | start > 870699:
cell_wall_organization_and_biogenesis_ (19.46)
| | | | | | | | | | | st̄art > 876625
| | | | | | | | | start <= 878072:
transcription_ (16.43)
| | | | | | | | | | start > 878072:
protein_modification_process_ (14.96)
| | | | | \ | \ start > 880365
| | | | | | | | | start <= 883824:
cytoskeleton_organization_and_biogenesis_ (18.12)
| | | | | | | | | start > 883824
| | | | | | | | | start <= 885792:
biological_process (18.5)
| | | | | | | | | | start > 885792:
DNA_metabolic_process_ (13.27)
| | | | | | start > 888970
| | | | | start <= 900751
| | | | | | | start <= 900190
| | | | | | | | | start <= 895957
| | | | | | | | | | start <= 892328:
RNA_metabolic_process_ (15.26)
| | | | | | | | | start > 892328:
protein_modification_process_ (17.69)
| | | | | | | | | start > 895957:
RNA_metabolic_process_ (18.78)
| | | | | | | | start > 900190:
protein_modification_process_ (22.63)
| | - | | \ | \overline{start > 900751}
| | | | | | | start <= 907214
| | | | | | | | | start <= 902040:
generation_of_precursor_metabolites_and_energy_ (21.35)
| | | | | | | | | start > 902040:
biological_process_(17.64)
| | | | | - | | | start > 907214
| | | | | | | | | start <= 911253: transcription_
(16.18)
| | | | | | | | | start > 911253:
RNA_metabolic_process_(25.99/12.19)
```



## Decision Tree Generated for Mitochondrial Chromosome:

## For Cellular Component

mitochondrion_(451.08)

## For Molecular Function

```
| start <= 61022
| | start <= 48201
| | | stop <= 25255
| | | | stop <= 18830
| | | | stop <= 16322
| | | | | | start <= 6546
| | | | | | start <= 731: molecular_function_ (8.76)
structural_molecule_activity_ (5.03)
| | | | | | start > 6546: molecular_function_ (19.53)
    stop > 16322: RNA_binding_ (12.48)
    stop > 18830
    | start <= 13818: hydrolase activity (40.54)
    | start > 13818: molecular_function_- (6.2)
    stop > 25255
    | start <= 28487
    start <= 13818: oxidoreductase_activity_(5.62)
```



## For Biological Process



```
| | | | | | start <= 13818: cellular_respiration_
(9.02/3.03)
| | | | | | start > 13818:
generation_of_precursor_metabolites_and_energy_ (24.65)
| | | | | start > 28487: translation_ (9.57)
| | | | start > 35373
| | | | start <= 36540
| | | | | | stop <= 42251: RNA_metabolic_process_
(17.57)
| | | | | | stop > 42251: electron_transport_
(18.14/7.95)
| | | | | start > 36540:
generation_of_precursor_metabolites_and_energy_ (6.14)
| start > 46723
| | start <= 72630
| | | stop <= 61729
| | | | start <= 48901: translation_ (20.25)
| | | | start > 48901: DNA_metaboli\overline{c}p\mathrm{ process_ (9.91)}
| | | stop > 61729: translation_ (195.92)
| | start > 72630
| | | start <= 85035
| | | | start <= 74495
| | | | | start <= 73758: cellular_respiration_
(22.53/9.96)
| | | | | start > 73758: biological_process_ (10.03)
| | | | start > 74495
| | | | | start <= 78533: translation_ (21.32)
| | | | | start > 78533
| | | | | start <= 79213
| | | | | | | neigh_num <= 1: cellular_respiration_
(8.28/2.69)
| | | | | | | neigh_num > 1: electron_transport_
(29.25/11.23)
| | | | | start > 79213: translation_ (13.98)
| | | start > 85035: RNA_metabolic_process_ (12.98)
```

