

University of Massachusetts Medical School

eScholarship@UMMS

---

GSBS Dissertations and Theses

Graduate School of Biomedical Sciences

---

2020-07-30


## Investigating Evolutionary Innovation in Yeast Heat Shock Protein 90

Pamela Cote-Hammarlof

*University of Massachusetts Medical School*

Let us know how access to this document benefits you.

Follow this and additional works at: [https://escholarship.umassmed.edu/gsbs\\_diss](https://escholarship.umassmed.edu/gsbs_diss)

 Part of the [Biochemistry Commons](#), [Environmental Microbiology and Microbial Ecology Commons](#), and the [Molecular Biology Commons](#)

---

### Repository Citation

Cote-Hammarlof P. (2020). Investigating Evolutionary Innovation in Yeast Heat Shock Protein 90. GSBS Dissertations and Theses. <https://doi.org/10.13028/tns2-1515>. Retrieved from [https://escholarship.umassmed.edu/gsbs\\_diss/1103](https://escholarship.umassmed.edu/gsbs_diss/1103)

Creative Commons License



This work is licensed under a [Creative Commons Attribution 4.0 License](#).

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in GSBS Dissertations and Theses by an authorized administrator of eScholarship@UMMS. For more information, please contact [Lisa.Palmer@umassmed.edu](mailto:Lisa.Palmer@umassmed.edu).

**INVESTIGATING EVOLUTIONARY INNOVATION IN YEAST HEAT  
SHOCK PROTEIN 90**

A Dissertation Presented

By

PAMELA COTE-HAMMARLOF

Submitted to the Faculty of the

University of Massachusetts Graduate School of Biomedical Sciences, Worcester

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

July 30<sup>th</sup>, 2020

BIOCHEMISTRY AND MOLECULAR PHARMACOLOGY

# INVESTIGATING EVOLUTIONARY INNOVATION IN YEAST HEAT

## SHOCK PROTEIN 90

A Dissertation Presented

By

PAMELA COTE-HAMMARLOF

The signatures of the Dissertation Defense Committee signifies completion and approval as to style and content of the Dissertation

---

Daniel N. A. Bolon, Ph.D., Thesis Advisor

---

Jennifer Benanti, Ph.D., Member of Committee

---

Mary Munson, Ph.D., Member of Committee

---

Nick Rind, Ph.D., Member of Committee

---

Daniel Jay, Ph.D., Member of Committee

The signature of the Chair of the Committee signifies that the written dissertation meets the requirements of the Dissertation Committee

---

Reid Gilmore, Ph.D., Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that the student has met all graduation requirements of the school.

---

Mary-Ellen Lane, Ph.D.,

Dean of the Graduate School of Biomedical Sciences

Biochemistry and Molecular Pharmacology  
(July 30, 2020)

## **Dedication**

I dedicate this dissertation to my husband, Tai Hammarlof, my son, Leonardo Cristobal Hammarlof and my grandmother, Phyliss Simonowitz. Thank you for your endless support, encouragement and love! Phyliss I know you are watching from above with a big smile on your face!

## Acknowledgement

First and foremost, I would like to give thanks to God because without him none of this is possible!

It truly takes a village to raise a doctoral student into a Doctor of Philosophy. As a result, there are many people I would like to thank. Firstly, I would like to thank my thesis mentor Dr. Daniel Bolon for giving me the opportunity to conduct my dissertation in your lab. Words cannot describe how truly thankful I am for your endless support, guidance, teaching and mentorship throughout my graduate training. Thank you for believing in me and allowing me to discover my potential! I would also like to thank the members of my Thesis Research Advisory Committee and Thesis Defense Committee, Dr. Reid Gilmore, Dr. Jennifer Benanti, Dr. Mary Munson, Dr. Nick Rhind, Dr. Konstantin Zeldovich and Dr. Daniel Jay for constructive criticism and support throughout my training and dissertation. I would also like to give a special thanks to Dr. Claudia Bank and Dr. Ines Fragata for their collaboration, constructive criticism, mathematical insight, knowledge and teaching of population genetics.

I would like to thank current and previous members of the Bolon lab. I would like to give a special thanks to Dr. Julia Flynn for your guidance, teaching, support and mentorship throughout my training in the lab! Thanks to Dr. Gily Nachum, Dr. Mohan Somasundaran, Dr. David Mavor, Neha Samant and Carl Hollins III, as well as previous members, Dr. Aneth Canale, Dr. Li Jiang, Dr. Parul Mishra, Ammeret Russow, Dr.

Jeffrey Boucher, Dr. Ben Roscoe and Dr. Ryan Hietpas for making the Bolon lab a great place to train as a graduate student. Thank you to everyone for being there to help me get through the ups and downs of scientific research and life! I would also like to thank Ryan Hietpas for your patience in teaching me the EMPIRIC method.

I would also like to thank the Biochemistry and Molecular Pharmacology Department for providing me with the supportive environment to promote my intellectual development as a student and scientist. Thank you to the BMP department administration members for being there whenever I needed assistance!

I would like to thank my informal mentors, Dr. Kendal Knight, Dr. Brian Lewis, Dr. Anthony Carruthers, Dean Mary Ellen Lane, and Dr. Teresita Padilla for your guidance and support!

Finally, I would like to thank my family! Thank you to my mom, Maryori Cote and my dad, Daniel Cote for encouraging my interests in science at such a young age and not denying me a telescope and microscope kit for Christmas! Thank you for standing behind me in my life endeavors and providing me with the hard work etiquette and determination to pursue my goals! Thank you for all that you have done to shape me into the person I am today! Thank you for your endless love and support! And thank you for taking care of Leonardo so that I can finish my graduate studies! Thank you to my sister, Amber Cote for making sure I never forget where I came from, keeping me in line, making me laugh and treating me like a Queen. Thank you to my other

mother, Carrie Sacco and my other father, Dr. Steve Scannell for your endless love, support and encouragement! Thank you for also taking such great care of Leonardo so that I can finish my dissertation.

Finally, I would like to thank two very important people in my life my husband, Tai Hammarlof and my son, Leonardo Hammarlof. Tai I dedicate this thesis to you because of your continuous love, support and patience throughout my ten plus years of education. Words cannot explain how eternally grateful I am for your encouragement and commitment throughout this lengthy and arduous process. Thank you for being by my side and for holding the fort down so that I could figure out what my interests are and allowing me to pursue those interests. Because of your commitment to us and my educational endeavors this thesis is as much mine as it is yours! Thank you from the bottom of my heart! Leonardo thank you for choosing me to be your mom. Thank you for giving me thesis breaks and teaching me how to have fun and how to look at the world around me with wonder again. I love you both eternally!

Thank you, Julia Flynn, and Aneth Canale for your edits and comments on my dissertation!

## Abstract

The Heat Shock Protein 90 (Hsp90) is an essential and highly conserved chaperone that facilitates the maturation of a wide array of client proteins, including many kinases. These clients in turn regulate a wide array of cellular processes, such as signal transduction, and transcriptional reprogramming. As a result, the activity of Hsp90 has the potential to influence physiology, which in turn may influence the ability to adapt to new environments. Previous studies using a deep mutational scanning approach, (EMPIRIC) identified multiple substitutions within a 9 amino acid substrate-binding loop of yeast Hsp90 that provides a growth advantage for yeast under elevated salinity conditions and costs of adaptation under alternate environments. These results demonstrate that genetic alterations to a small region of Hsp90 can contribute to evolutionary change and promote adaptation to specific environments. However, because Hsp90 is a large, highly dynamic and multi-functional protein the adaptive potential and evolutionary constraints of Hsp90 across diverse environments requires further investigation.

In this dissertation I used a modified version of EMPIRIC to examine the impact of environmental stress on the adaptive potential, costs and evolutionary constraints for a 118 amino acid functional region of the middle domain of yeast Hsp90 under endogenous expression levels and the entire Hsp90 protein sequence under low expression levels. Endogenous Hsp90 expression levels were used to observe how environment may affect Hsp90 mutant fitness effects in nature, while low expression levels were used as a sensitive readout of Hsp90 function and fitness. In



general, I found that mutations within the middle domain of Hsp90 have similar fitness effects across many environments, whereas, under low Hsp90 expression I found that the fitness effects of Hsp90 mutants differed between environments. Under individual conditions multiple variants provided a growth advantage, however these variants exhibited growth defects in other environments, indicating costs of adaptation. When comparing experimental results to 261 extant eukaryotic sequences I find that natural variants of Hsp90 support growth in all environments. I identified protein regions that are enriched in beneficial, deleterious and costly mutations that coincides with residues involved in co-chaperone-client-binding interactions, stabilization of Hsp90 client-binding interfaces, stabilization of Hsp90 interdomains and ATPase chaperone activity.

In summary, this thesis uncovers the adaptive potential, costs of adaptation and evolutionary constraints of Hsp90 mutations across several environments. These results complement and extend known structural and functional information, highlighting potential adaptive mechanisms. Furthermore, this work elucidates the impact environment can have on shaping Hsp90 evolution and suggests that fluctuating environments may have played a role in the long-term evolution of Hsp90.

## Table of Contents

Title Page	II
Dedication	III
Acknowledgement	IV
Abstract	VII
Table of Contents	IX
List of Tables	XIII
List of Figures	XIV
List of Copyrighted Materials and Third Party Copyrighted Materials	XXI
List of Abbreviations	XXII
<b>Chapter I Introduction</b>	<b>1</b>
<hr/>	
Adaptation to Novel Environment via Natural Selection	1
Mechanistic Insight into How Natural Selection Can Promote Adaptation	2
Mendelian Inheritance and Mutation	2
Molecular Mechanisms of Adaptive Evolution	6
In Vitro Experimental Evolution	9
<i>In Vitro</i> Selection and Directed evolution	14
Nucleic Acid Sequencing Technology	18

	X
Next Generation Sequencing Technology	20
Mutational Scanning Approaches to Reveal Protein Fitness Landscapes	21
Fitness	28
Distribution of Mutant Fitness Effects	31
The Heat Shock Protein 90	34
Heat Shock Protein 90 Function and Structure	34
Hsp90 and Adaptation to Stressful Environments	40
Hsp90 Facilitates the Rapid Evolution of Novel Traits	43
Thesis Scope	44
<b>Chapter II: The Adaptive Potential of the Middle Domain of Yeast Hsp90</b>	<b>49</b>
<hr/>	
Abstract	50
Introduction	51
Results and Discussion	55
DFEs Across All Environments Show Many Wild-Type Like Mutations	55
Correlations Across Conditions Reveal That Diamide Stands Out with Respect to	
Mutational Effects	57
Beneficial Mutations are Present Across All Environments and Are Enriched in A	
Region of Hsp90 That is Implicated in Stabilization of Interdomains and Client	

	XI
Binding Interfaces	60
The Proportion of Deleterious, Wild-Type Like, and Beneficial Mutations Varies Greatly Along the Protein Sequence	66
Few Mutations Show Costs of Adaptation	67
Conclusions	70
Materials and Methods	71
Generating Point Mutations	71
Constructing Barcoded Libraries	72
Associating Barcodes to Mutants	73
Yeast Competitions	74
Sequencing of Competition Samples	76
Estimation of Selection Coefficients	77
Natural Variants	78
Costs of Adaptation	78
Acknowledgements	79
<b>Chapter III: Comprehensive Fitness Maps of Hsp90 Show Widespread Environmental Dependence</b>	<b>102</b>
<hr/>	
Abstract	103
Introduction	104

Results	107
Impact of Stress Conditions on Mutational Sensitivity of Hsp90	113
Structural Analyses of Environmental Responsive Positions	118
Constraint of Mutational Sensitivity at High Temperature	121
Hsp90 Potential for Adaptation to Environmental Stress	126
Natural Selection Favors Hsp90 Variants That are Robust to Environment	130
Discussion	133
Limitations	135
Relationship to Prior Work	136
Materials and Methods	138
Generating Mutant Libraries	138
Barcode Association of Library Variants	139
Bulk Growth Competitions	140
DNA Preparation and Sequencing	141
Analysis of Bulk Competition Sequencing Data	142
Determination of Selection Coefficient	142
Structural Analysis	144
Random Simulations to Assess Clustering of Mutations	144
Yeast Growth Analysis	144
Analysis of Hsp90 Expression by Western Blot	145

	XIII
Natural Variation in Hsp90 Sequence	146
Acknowledgements	146
<b>Chapter IV: General Discussion</b>	<b>172</b>
<hr/>	
Summary	172
Most Mutations Exhibit Neutral and Deleterious Fitness Effects Across Environments	175
The Impact of Stress on Average Mutant Fitness Effects and Selection of Hsp90 Mutations	178
The Adaptive Potential of Hsp90 Across Environmental Conditions	180
Diamide and 37°C Places Distinct Selection Pressures on Hsp90 Mutations	184
Concluding Remarks and Future Studies	188
Bibliography	191

### **List of Tables**

Supplementary Table 2.1: Number of beneficial, neutral and deleterious mutations across environments, and replicates	99
Supplementary Table 2.2: Pearson correlation of fitness effects across replicates and environments.	100
Supplementary Table 2.3: Percentage (and number) of beneficials and their cost across environments	101
Table 3.1: Key resource table	171

## List of Figures

Figure 1.1: Yeast Hsp90 structure	37
Figure 1.2: Current model of Hsp90 mediated client maturation.	38
Figure 2.1: Middle domain of Hsp90	54
Figure 2.2: Histogram of correlations of fitness effects among replicates and environments reveals that diamide stands out as different	59
Figure 2.3: Hotspots of beneficial and deleterious mutations in the middle domain do not coincide with known client binding loop	65
Figure 2.4: Histogram of proportions of mutations that display costs of adaptation illustrates that costs of adaptation are generally rare, and most likely if comparisons involve diamide environments	68
Figure 2.5: A hotspot of costs of adaptation is located at positions 381-391	69
Supplementary Figure 2.1: Doubling times for WT Hsp90 yeast growth under all environments	80
Supplementary Figure 2.2: Distribution of Fitness Effects (DFE) across environments	81
Supplementary Figure 2.3: Heatmaps of mutant fitness effects	82
Supplementary Figure 2.4: Distribution of selection coefficients of mutations found in natural environments vs all mutations studied here	88

	XV
Supplementary Figure 2.5: Fraction of mutations that are beneficial, deleterious or neutral from our study and occurring in natural populations	89
Supplementary Figure 2.6: Correlation between replicates for each environment	90
Supplementary Figure 2.7: Correlation between environments in 10 amino acid regions	91
Supplementary Figure 2.8: Proportions of deleterious, beneficial and wild-type like mutations along the middle domain of Hsp90	92
Supplementary Figure 2.9: Summary of the proportions of beneficial, deleterious and neutral mutations in each environment/replicate	93
Supplementary Figure 2.10: Distribution of fitness effects at position 364	94
Supplementary Figure 2.11: Cost of adaptation in different environments	95
Supplementary Figure 2.12: Barcoding mutant library strategy	96
Supplementary Figure 2.13: Average number of barcodes per mutant for each 9 amino acid window shows an average of ~20 barcodes associated with each variant	97



	XVI
Supplementary Figure 2.14: Strategy how to associate barcodes to mutants	98
Figure 3.1: Approach to determine protein fitness maps of Hsp90	109
Figure 3.2: Impact of environmental stresses on yeast growth rates and selection on Hsp90 sequence	115
Figure 3.3: Environmental stresses place distinct selection pressures on Hsp90	119
Figure 3.4: Abundance and mechanism of temperature-sensitive ( <i>ts</i> ) mutations in Hsp90	123
Figure 3.5: Beneficial variants in diamide and elevated temperature conditions	127

Figure 3.6: Experimental growth effects of natural amino acid variants of Hsp90	132
Supplementary Figure 3.1: Selection coefficients for wild-type synonyms (green) and stops (red) at each position of Hsp90 for both replicates of standard conditions	147
Supplementary Figure 3.2: Measurement of selection coefficients for positions 2–220 in this study correlated strongly ( $R^2 = 0.87$ ) with estimates of the Hsp90 N-domain in a previous study (Mishra et al., 2016), indicating that biological replicates show high reproducibility	148
Supplementary Figure 3.3: Analysis of variation in stop codon selection coefficients	149
Supplementary Figure 3.4: Heatmap representation of the selection coefficients observed for single amino acid changes across amino acids 2–709 of Hsp90 in standard (30°C) conditions in replicate 1	150
Supplementary Figure 3.5: Correlation of mutational sensitivity with distance to ATP	151
Supplementary Figure 3.6: Heatmap representation of the fitness map observed for single amino acid changes of Hsp90 in nitrogen depletion	152

Supplementary Figure 3.7: Heatmap representation of the fitness map observed for single amino acid changes of Hsp90 in salt	153
Supplementary Figure 3.8: Heatmap representation of the fitness map observed for single amino acid changes of Hsp90 in ethanol	154
Supplementary Figure 3.9: Heatmap representation of the fitness map observed for single amino acid changes of Hsp90 in diamide	155
Supplementary Figure 3.10: Heatmap representation of the fitness map observed for single amino acid changes of Hsp90 at 37C	156
Supplementary Figure 3.11: Distribution of selection coefficients for non-synonymous mutations (black), wild-type synonyms (green), and stops (red) in each environmental condition	157
Supplementary Figure 3.12: Distribution of selection coefficients in each environmental condition	158

Supplementary Figure 3.13: Analysis of variation in wild-type synonym selection coefficient	159
Supplementary Figure 3.14: Distribution of the difference between selection coefficients of each mutation in each stress condition and the same mutation in standard conditions	161
Supplementary Figure 3.15: Heatmap representation of the average selection coefficient (s) at each position in each environmental condition relative to the average selection coefficient at the same position in standard conditions	162
Supplementary Figure 3.16: Environmentally responsive Hsp90 positions are enriched in binding contacts	163
Supplementary Figure 3.17: Validation of beneficial mutants at 37°C	164
Supplementary Figure 3.18: Distribution of the number of beneficial mutations at the same position in standard, nitrogen depletion, salt, and ethanol conditions	165

Supplementary Figure 3.19: Selection coefficients in standard conditions for all wild-type-like mutations at 37°C and in diamide	166
Supplementary Figure 3.20: Synonymous mutations at the beginning of Hsp90 have strong beneficial growth effects	167
Supplementary Figure 3.21: Hsp90 expression in the Hsp90 shutoff yeast strain harboring either wild-type Hsp90 under the constitutive ADH promoter or a null plasmid (no insert)	168
Supplementary Figure 3.22: Heatmap of Spearman's rank correlation coefficients ( $r$ ) between molecular features (rows) and selection coefficients per mutation or per position (average selection coefficient for all amino acids at the position) for each environment (columns)	169
Supplementary Figure 3.23: Validation of yeast Hsp90-shutoff strain	170

### List of Copyrighted Materials Produced by the Author

Chapter II represents work previously published previously and is presented in accordance with copyright law.

*Pamela A. Cote-Hammarlof*<sup>1\*</sup>, *Inês Fragata*<sup>2\*§</sup>, *Julia Flynn*<sup>1</sup>, *David Mavor*<sup>1</sup>, *Konstantin B. Zeldovich*<sup>1</sup>, *Claudia Bank*<sup>2,3#</sup>, and *Daniel N.A. Bolon*<sup>1#</sup>. **The adaptive potential of the middle domain of yeast Hsp90.** (\*equal contribution; # co-corresponding author) *Molecular Biology and Evolution*, msaa211, <https://doi.org/10.1093/molbev/msaa211>

Chapter III represents work previously published previously and is presented in accordance with copyright law.

*Julia M. Flynn*<sup>1</sup>, *Ammeret Rossouw*<sup>1</sup>, *Pamela A. Cote-Hammarlof*<sup>1</sup>, *Ines Fragata*<sup>2</sup>, *David Mavor*<sup>1</sup>, *Carl Hollins III*<sup>1</sup>, *Claudia Bank*<sup>2</sup>, and *Daniel N.A. Bolon*<sup>\*</sup>. 2020. **Comprehensive fitness maps of Hsp90 show widespread environmental dependence.** *Elife* 9: e53810. Published 2020 Mar 4. doi:10.7554/eLife.53810

### Third Party Copyrighted Material

The following figure was reproduced from journal: Permission required

Figure	Number	Publisher License Number
Figure 1.2	Springer Nature	4906010478105

The following figures were reproduced from journals: No permission required

Figure Number	Publisher
Figure 1.1	Plos ONE

**List of Abbreviations**

aa – Amino Acid

ADP - Adenosine Diphosphate

ATP - Adenosine Triphosphate

C-terminal - Carboxy-Terminal

dNTP- Deoxynucleotides

ddNTP - Dideoxynucleotide Triphosphate

DE - Directed Evolution

DFE - Distribution of Fitness Effects

DMS- Deep Mutational Scanning

DNA - Deoxyribonucleic Acid

EMPIRIC - Exceedingly Methodical and Parallel Investigation of Randomized Individual Codons

*E. coli- Escherichia coli*

FACS - Fluorescence-Activated Cell Sorting

FGM - Fisher's Geometric Model

GERP- Genomic Evolutionary Rate Profiling

GFP - Green Fluorescent Protein

GR- Glucocorticoid Receptor

hGH- Human Growth Hormone

hGHp- Human Growth Hormone partner

HIV - Human Immunodeficiency Virus

Hsp90 - Heat Shock Protein 90

M-domain- Middle domain

N-terminal - Amino-Terminal

NGS- Next Generation Sequencing

PCR - Polymerase Chain Reaction

s- Selection Coefficient

*S. cerevisiae* - *Saccharomyces cerevisiae*

SELEX - Systematic Evolution of Ligands by Exponential Enrichment

SERF - Self-Encoded Removable Fragment

SIFT- Sorting Intolerant from Tolerant

ts- Temperature sensitive

wt - Wild Type



## Chapter I: Introduction

### Adaptation to Novel Environments via Natural Selection

Individuals are constantly subjected to novel and fluctuating environments that can impose selective pressures on an individual's fitness such as, abiotic stress (temperature, pH, salinity), limited food resources and predation. However, many organisms can survive, reproduce and thrive when subjected to these selective pressures in nature, resulting in the survival of diverse species within distinct environmental niches. As a result of this biological phenomenon, understanding the mechanism of how organisms can adapt to novel and stressful environments has been and continues to be a central question.

Since the classical era, Greek philosophers have contemplated theories to describe and understand how organisms can survive and persist when presented with selective pressures in nature. The physician, Hippocrates the II expressed the idea that an organism can pass on to its offspring physical traits that promote adaptation to specific environments (Zirkle 1941). At the same time, the philosopher Empedocles postulated a theory of natural selection, whereby reproductive fitness, random chance and survival of the fittest promotes adaptation to environmental change (Zirkle 1941). The contemplation of these theories during this time were contradictory to the common belief that supernatural forces were dictators of life and death. It was not until the 19<sup>th</sup> century that ideas related to natural selection were revisited and popularized by Lamarck, Darwin and Wallace to provide a theoretical mechanism for adaptive evolution in nature. The zoologist, Jean Baptiste Lamarck postulated that the environment induces an adaptive phenotypic change that results in an environment specific

phenotype that is subsequently inherited (Lamarck 1914). His theory was later rejected due to the combined works of Alfred Wallace and Charles Darwin on Natural Selection. Charles Darwin presented his famous theory of evolution by natural selection in his book titled "The Origin of Species by Natural Selection". Darwin coined "natural selection as the mechanism that causes evolutionary change via differences in an individual's phenotype that helps individuals adapt to their environment and reproduce" (Darwin 1859). The subsequent heritability of these characteristics further promotes survival of populations and continuance of a species should the environment change (Darwin 1859), whereas, individuals who are considered less fit cannot reproduce and go extinct (Darwin 1859). Darwin synthesized his theory of natural selection through observations of natural populations of animals on his 4-year voyage of the Beagle along with research of Robert Malthus's topic on the struggles of existence in the "Principles of Population" and letters from Alfred Wallace's theory on natural selection. Wallace independently synthesized his theory of natural selection from his own observations of natural populations in the Amazon (Wallace 2013). Together the work of Darwin and Wallace provided a theoretical model for adaptive evolution, which has become a central tenet of biology. However, the physiological mechanism by which natural selection can act upon to promote adaptive evolution remained elusive and controversial.

### **Mechanistic Insight into How Natural Selection Can Promote Adaptation**

#### **Mendelian Inheritance and Mutation**

Around the same time that Darwin proposed his theory of Natural Selection, Gregor Mendel discovered the mechanism of inheritance through investigations of

hybridization in pea plants. He demonstrated that the inheritance of phenotypic traits was an isolated process, that was discontinuous and could be quantitatively predicted (Mendel 1866). He also theorized that this mechanism of inheritance was compatible with Darwin's theory of natural selection based on his own studies and interpretations of Darwin's theory. Unfortunately, his work was dismissed at the time because Darwin and others believed that adaptive traits were inherited via a blending of parental traits (Darwin 1859). It was not until the following century that Mendel's work would be fully appreciated.

The beginning of 20<sup>th</sup> century brought forth a large body of important work and discovery in the fields of Genetics and the quantitative field of Population Genetics, which ultimately unified Mendelian Inheritance with Darwin's Theory of Natural Selection. At the turn of the century the geneticists, Correns (Correns 1950; Corcos and Monaghan 1987), Tschermak (Tschermak 1950) and De Vries (de Vries 1950) rediscovered Mendel's work on the inheritance of phenotypic traits in pea plants. De Vries was also the first to discuss the idea of mutations as the cause of observed phenotypic differences (de Vries 1950). However, it was the biologist, William Bateson who further popularized Mendel's mechanisms of biological inheritance as the mechanism that coupled to mutation and natural selection can result in phenotypic differences among organism (Bateson 1894). Most importantly, he invented the terms "genetics" to describe the study of heredity and "allele" as the agent of inheritance (Gillham 2001). Together, Correns, Tschermak, De Vries and Bateson helped spread Mendel's concept of inheritance and principles that has led to our current understanding of the relationship among inheritance, mutation and natural selection. Their work also

laid down the foundation for later efforts by evolutionary biologists Ronald Fisher, J.B.S. Haldane and Sewall Wright on population genetics.

Fisher, Haldane and Wright established the field of population genetics, which uses quantitative models to demonstrate how Mendelian genetics and gene frequency change is compatible with natural selection and adaptation in populations (Haldane 1927; Fisher 1931; Wright 1932; Haldane 1959). Furthermore, Fisher and Wright were pioneers in the development of fitness landscape models that could depict how individuals and organisms could sample a multi-dimensional phenotypic space during natural selection and adaptation to specific environments. Specifically, Ronald Fisher's work outlined an adaptive evolutionary model, "Fisher's Geometric model" (FGM) in his book, "The Genetical Theory of Natural Selection". Fisher's model allows one to physically map the effects of phenotypic mutations to fitness and form quantitative predictions of the fraction of available beneficial, "adaptive" mutations based on the population's initial level of fitness (Fisher 1931). Specifically, Fisher's model promotes the characterization of an initial population with a given fitness distanced from an 'optimum' fitness as a 'point' in a multidimensional space (Fisher 1931). Mutations that arise are characterized as 'vectors' that are generated in random directions and magnitudes away from the optimum in the multidimensional space (Fisher 1931). Vectors that bring the initial high fitness population furthest away from the optimum are considered deleterious, whereas, those that bring the initial populations closest to the optimum are beneficial (Fisher 1931). Most importantly, his model helps to explain how the distribution of spontaneous mutant fitness effects and effect sizes in a fitness landscape can contribute to continuous variation of traits (Fisher 1931; Waxman 2006).

Around the same time, Sewall Wright introduced a similar yet distinct concept of adaptive landscapes to model the relationship between genotype-phenotype and fitness in his book, "The Roles of Mutation, Inbreeding, Crossbreeding, and Selection in Evolution" (Wright 1932). Wright's concept was based on initial populations in a landscape climbing to nearest peaks of adaptive fitness as a result of natural selection acting upon mutations (Wright 1932). His work on adaptive landscapes showed that genetic drift, whereby random sampling of genetic variants causes a change in the frequency of existing variants in a population, resulting in either a reduction or fixation of alleles in populations (Wright 1932). J. B. S. Haldane derived mathematical expressions to estimate the direction and rate of gene frequency change in natural populations in his series of papers on "A Mathematical Theory of Natural and Artificial Selection". These mathematical models help to promote the analysis of the interaction of mutation with natural selection and demonstrated that natural selection acts faster than what was previously thought (Haldane 1927). Together they provided critical quantitative methods that could predict, measure and test natural selection. Additionally, they provided key concepts in the fields of genetics and evolution, which Theodosius Dobzhansky further popularized along with experimental evidence in his book, "Genetics and the Origin of Species" (Dobzhansky 1937). Most importantly, Dobzhansky summarization combined Mendelian Genetics with Darwinian Evolution, which served as the catalyst for the modern synthesis of evolutionary biology (Ayala and Fitch 1997).

The significant contributions in population genetics and experimental genetics helped to provide ways to predict adaptive evolution and further unify the fields of

Mendelian genetics and Darwinian evolution via natural selection. Moreover, they uncovered important concepts in the field that lead to the conception of Julian Huxley's modern synthesis of evolutionary biology, which are a set of principles that serve as a framework to understand evolutionary mechanisms in all biological fields. Huxley summarized his synthesis in his 1942 book, "Evolution the Modern Synthesis" and emphasized that the origin of adaptive evolution is due to natural selection acting on inherited genetic variability that arises via random changes in genetic material, otherwise known as 'mutations' (Huxley 1944; Charlesworth et al. 2017). He also brought forth the notion that these mutations arise regardless of their effects on fitness, whether mutations are advantageous or disadvantageous (Huxley 1944; Charlesworth et al. 2017). While Huxley's modern synthesis provided the framework for our current understanding of evolutionary mechanisms it was not until the 1950's that we began to understand how genes and the instructions for them are held within organisms and passed on from generation to generation. Most importantly, how mutations can arise within organisms, serve as the catalyst for evolutionary change and passed on to subsequent generations.

### **Molecular Mechanisms of Adaptive Evolution**

The middle of the 20<sup>th</sup> century was an important time in scientific history because of the important contributions in the fields of Molecular Biology that provided insight into the molecular aspects of evolution and further unification of inheritance with natural selection. In 1953 the biologist, James Watson and the physicist, Francis Crick reported a model of the deoxyribonucleic acid (DNA) double helix based on Rosalind Franklin's crystallographic work on DNA combined with the mathematics of a helix

transform (Watson and Crick 1953). This discovery was crucial to science because it precisely and correctly outlined the current molecular structure of DNA, as an anti-parallel double stranded polymer composed of repeated units of nucleotides (Adenine, Guanine, Cytosine and Thymine). Most importantly, they uncovered that only two specific base-pairing combinations exist, Adenine-Thymine and Guanine-Cytosine, which they proposed could serve as a copying mechanism for genetic material and mispairings between these bases could cause a mutation during replication (Watson and Crick 1953). Around the same time Matthew Meselson and Franklin Stahl's famous experimental studies of bacterial replication and phage infection in *E. coli* confirmed that DNA can carry and transmit hereditary information and is self-replicating, which further supported Crick and Watson's hypothesis (Meselson and Stahl 1958). Later genetic experiments by Crick, Barnett, Brenner and Watts-Tobin on bacteriophage T4 mutants and analysis of tobacco mosaic virus mutants by Wittmann, Tsugita and Fraenkel-Conrat demonstrated through the discovery of frameshift mutations that the genetic code for proteins is degenerate, and composed of non-overlapping triplets of bases, "codons" derived from part of the nucleic acid (Crick 1970). This discovery laid the foundation for future work by Khorana, Holley and Nirenberg on deciphering the genetic code and describing how the code synthesizes proteins from DNA which is transcribed into RNA and translated into protein (Nirenberg and Matthaei 1961; Holley et al. 1965; Nirenberg et al. 1965; Kresge et al. 2009) . Together the extraordinary discoveries during this time highlighted the molecular mechanisms that promotes the carry and transmission of hereditary material and the flow of genetic information within an organism, "from nucleic acid to nucleic acid and nucleic acid to protein but not protein to

nucleic acid” , which Crick coined the ‘Central Dogma of Biology’ (Crick 1958; Cobb 2017).

The discovery of DNA and the molecular mechanisms that enable the transmission of hereditary material provided new perspectives on the genetic basis of adaptation, resulting in the study of molecular mechanisms of variation and their evolutionary consequences. The development of “Fingerprinting” techniques, such as gel electrophoresis, immune assays and paper chromatography allowed for the study of evolutionary mechanisms at both the protein and sequence level. For example, the advent of protein sequencing technology allowed for sequence comparisons between homologous protein sequences. By directly comparing sequences one could then estimate divergence times between two sequences via determining the mutation rate per generation and the number of nucleotide differences between two sequences, coined the “molecular evolutionary clock” (Zuckerkandl and Pauling 1965). These estimates helped to uncover at what time in history evolutionary changes occurred. Other studies looked at the genetic variation in humans and animals to quantify the relationship between species and within populations, which helped to uncover other important evolutionary concepts. Specifically, studies of genetic variation within dozens of *Drosophila pseudoobscura* loci found a large proportion of loci were polymorphic (Lewontin and Hubby 1966), which they attributed to either balancing selection, which can maintain various alleles in a species through natural selection mechanisms or neutral mutations that have no effect on fitness as the explanation for their observations. Most importantly, their indication that neutral mutations could account for



the high levels of variability within species served as the basis for future work on the “Neutral Theory of Evolution” by Motoo Kimura (Kimura 1970; Kimura 1983).

### ***In Vitro* Experimental Evolution**

Even though we gained a further understanding of the molecular mechanisms of adaptation and population genetic models provided important evolutionary concepts that can impact evolutionary processes and patterns there remained gaps in knowledge between theoretical models and observed phenomena. Furthermore, addressing questions in population genetics and what occurs in nature is difficult due to differences in environments, taxa, life span and generation times (Bailey and Bataillon 2016). As a result of this came the rise of *in vitro* experimental evolution, which allows for repeated phenotypic and genetic studies of populations from a common ancestor in controlled environments on shorter timescales than what is observed in nature. These studies allow one to get a glimpse of what happens in nature. For example, one of the earliest controlled experimental evolution experiments described was performed by William Ballinger from 1880-1886. Ballinger created an incubator to look at the effects of increased temperatures on the survival of protists and cultivated strains that could withstand temperatures that were initially lethal to the protists (Lenski 2017). His work was one of the first to observe evolution in action and provide evidence of Darwinian adaptation. However, Ballinger could not identify or describe the mechanism that resulted in the thermal adaptive mutants. It was not until the early 20<sup>th</sup> century, that controlled laboratory studies of fruit flies provided ways in which to study adaptation and identify what factors are involved in this process. For example, Theodosius Dobzhansky

studies of experimentally bred populations of *Drosophila pseudoobscura* allowed for him to observe extensive genetic variation within species, and provide further evidence to support the notion that mutations allow populations to evolve rapidly to changing environments (Dobzhansky 1937; Ayala and Fitch 1997). Even though *Drosophila* and other unicellular organisms were used as model organisms to study experimental evolution, bacteria were initially excluded as model organisms because at the time it was thought that bacteria did not have genes (Huxley 1944). It was not until the 1940's that Max Delbrück and Salvador Luria used bacteria in their "fluctuation test" to estimate mutation rates and test whether mutations occur before or after selection (Luria and Delbrück 1943). They showed that mutation and selection were unique processes in *Escherichia coli* grown in the presence and absence of T4 viral phage, where mutations arise in the absence of selection pressure instead of in response to selection pressure imposed by T4 viral phage (Luria and Delbrück 1943). This discovery was critical because it showed that Darwin's theory of natural selection acting upon random mutations applied to bacteria. Others began to use bacteria to address similar questions, for example Aaron Novick and Leo Szilard worked on ways to estimate the rate of mutations by measuring the rate of spontaneous accumulation of mutations in *E. coli* populations grown in a chemostat, a device that Novick and Szilard created that could keep a population of bacteria growing at a reduced growth rate for an undefined time (Novick and Szilard 1950). Through their studies they observed a drastic decline in initial mutant populations followed by the linear accumulation of adaptive mutants that displaced the parental strain, which provided experimental evidence in support of their estimates of mutation rates (Novick and Szilard 1950). Together the work of Luria,

Delbruck, Novick and Szilard showed the scientific community that bacteria were excellent models to study adaptation with the advantages of shorter generation times and large population sizes and the ease with which environments can be controlled and manipulated. As a result of this was the birth of the field of microbial experimental evolution.

In the late 1980's the evolutionary biologist, Richard Lenski pioneered a long-term experimental evolution experiment with 12 identical populations of *E. coli*. from two ancestral strains (Lenski et al. 1991). These strains have been continuously growing in a selective environment of glucose limiting high citrate conditions and diluted in 24-hour cycles up to 73,000 generations, until a pause on March 9th 2020 due to the COVID pandemic (Lenski 2020). Through phenotypic observation and tracking of genetic variation in all populations throughout the course of this experiment Lenski has been able to investigate various aspects of adaptation including: the dynamics of adaptation, such as, drift and repeatability (Lenski and Travisano 1994); estimates of temporal patterns of evolution, otherwise known as the "molecular clock" (Wielgoss et al. 2013); and epistasis (Khan et al. 2011). For example, at ~ 2,000 generations he observed populations with large increases in growth rate due to the step- wise rise of spontaneous beneficial mutations in the genetic background. Over time these mutants became less beneficial, which he attributed to the rise of competition between different beneficial lineages (termed clonal interference) and the small effects these beneficial mutations produce in more-fit compared to less-fit backgrounds (termed diminishing returns epistasis) (Lenski et al. 1991; Lenski and Travisano 1994). However, at ~ 50,000 generations 70% of the populations grew faster than the ancestral population

and fitting the fitness trajectories of these beneficial mutations to theoretical models showed that both adaptation and divergence can occur indefinitely (Wiser et al. 2013; Lenski et al. 2015). In addition, by incorporating clonal interference and epistasis into a theoretical model of changes in mean fitness over time for large asexual populations they were able to generate similar results to what they observed empirically (Wiser et al. 2013; Lenski et al. 2015). The most fascinating discovery occurred at ~ 30,000 generations, when Lenski observed a population that had evolved the ability to use citrate as a carbon source under glucose limiting environments (Blount et al. 2008). Further genomic analysis of these mutants determined that a duplication of the *citT* operon along with disruptions in the regulation of this operon and at least one single nucleotide polymorphism in the genetic background contributed to this adaptive phenotype (Blount et al. 2012). Collectively, the long-term experimental evolution experiment in *E. coli* has allowed for the continued observation of adaptation, identification of important mechanistic factors that affects fitness and adaptation and a clearer understanding of the relationship between natural selection and adaptation in *E. coli*. Furthermore, these experiments served as a stepping stone for the use of other microbial systems in the exploration of experimental evolution.

*Saccharomyces cerevisiae* and the fission yeast, *Schizosaccharomyces pombe* have become popular models to study adaptation for many reasons. Yeast contain similar cellular machinery to higher eukaryotes, many yeast proteins have human homologues, many mechanisms and pathways are conserved within humans, their genomes are easy to manipulate and their different reproductive states (diploid or haploid) are easy to control. Therefore, one can use yeast to address an array of

questions related to adaptation at the genetic, and phenotypic level that would provide a glimpse into adaptive evolutionary mechanisms in higher eukaryotes. Examples of evolutionary mechanisms that have been investigated in yeast include: how often and when do mutations occur in DNA (Lang and Murray 2008); the roles of clonal interference; allele frequency changes (genetic hitchhiking ) and evolutionary trade-offs in function plays in the repeatability of adaptation (Kvitek and Sherlock 2013; Lang et al. 2013); evolution of multicellularity (Ratcliff et al. 2012); and sex (Hill and Otto 2007). Yeast have been also used extensively to investigate the effects of various environmental perturbations on yeast adaptation at both the genetic, expression and phenotypic level to provide an understanding of how adaptation to novel and stressful environments occurs mechanistically. For example, Dhar et al., studied gene expression and sequence changes in *S. cerevisiae* grown under continuous high salt conditions for 300 generations and found evidence for an increase in genome size and modest changes in the expression of several genes as modes of adaptation (Dhar et al. 2011). Others investigated patterns of gene expression changes during adaptation to stressful environments including, nutrient limiting conditions (Ferea et al. 1999; Gasch et al. 2000), chemical and heat perturbations (Gasch et al. 2000) and discovered large changes in the expression of certain subsets of genes including environmental stress response, heat shock and metabolic genes along with identification of gene expression regulators (Ferea et al. 1999; Gasch et al. 2000). More recently, Gorter et al., looked at the effect of gradually increasing or constant high concentrations of the heavy metals cadmium, nickel, and zinc on genomic evolution in *S. cerevisiae* and found that adaptation to these metals required a combination of small nucleotide polymorphisms,

small indels, and whole-genome duplications to occur (Gorter et al. 2017). Together, experimental evolution studies in yeast have shed light on adaptive evolutionary mechanisms in higher eukaryotes

Thus, experimental evolution allows for the controlled study of organismal and population adaptation to new environmental conditions by natural selection. Most importantly, experimental evolution promotes the analysis of underlying factors that contribute to adaptive evolution including, gene expression changes within cells and the proportion of beneficial mutations that occur in a population. Both methods have provided valuable insight into adaptive mechanisms and evolution. However, experimental evolution has some limitations that prevent one from gaining a complete understanding of evolutionary mechanisms. For example, measurement of mutant accumulation during experimental evolution is limiting because it only captures relatively fit or beneficial mutations that remain in the population over many generations. Therefore, these experiments do not provide information on the magnitude of these beneficial mutations on fitness or the fitness effects of all mutations.

### ***In Vitro* Selection and Directed Evolution**

Mutagenesis approaches are based on the logic that perturbations to the protein sequence results in functional consequences that can impact an individual's phenotype and or fitness. As a result, mutagenesis approaches are useful to molecular geneticists and evolutionary biologists because these approaches enable investigation of how mutations can impact various biological phenomena. Traditional mutagenesis approaches incorporate a forward genetics approach, whereby random spontaneous

mutations accumulate and produce specific phenotypes that are screened, followed by mapping, cloning, sequencing and finally annotating the mutant gene (Botstein and Shortle 1985). One of the first mutant accumulation experiments studied the spontaneous mutation rate of polygenic mutations that control viability and their impact on the genetic structure in natural populations of *Drosophila melanogaster* (Mukai 1964). Using this approach Mukai discovered that large rates of spontaneous polygenic mutations may contribute to large observations of genetic variation in *Drosophila* (Mukai 1964). Similar yet distinct approaches applied x-rays to induce spontaneous mutations within genes in *Drosophila* (Muller 1927) and deduced that X rays induced lethal mutations and that egg and sperm cells are more susceptible to genetic mutations (Muller 1927). While others used chemical mutagens to induce random mutations within organisms such as, *Aerobacter aerogenes* to study the evolution of catabolic pathways in bacteria (Lerner et al. 1964). Together these approaches have provided insight into evolutionary mechanisms, however these approaches are limiting because they rely on the occurrence of random spontaneous mutations that affect phenotype, which results in a low probability of investigation of mutations of interest and limitations in studying the full spectrum of mutations. The discovery of mobile genetic elements, termed transposons (McClintock 1950) resulted in the generation of transposon mutagenesis, which is an improved targeted forward genetics approach that results in specific insertion of one mutation per genome, followed by screening and transposon tagging to identify and clone genes of interest (Ruvkun and Ausubel 1981). Ruvkun et al., used this technique to study the symbiotic nitrogen fixation genes of *Rhizobium meliloti* and construct a physical maps of a subset of these genes (Ruvkun and Ausubel 1981).

It was not until the 1990's that advances in recombinant DNA technologies could produce mutations within distinct sites of a gene or randomly throughout the entire gene. Approaches that can produce mutations within specific regions of a gene include, polymerase chain reaction (PCR) combined with synthetic oligonucleotides and thermocycling to induce single mutations in a gene (Schochetman et al. 1988), or cassette mutagenesis, which is a cassette composed of annealed oligonucleotides with single mutations that is ligated into a region of a gene where the wild-type sequence has been previously removed (Wells et al. 1985). Whereas, approaches like error prone PCR uses a polymerase without DNA proof-reading ability to enable the generation of random mutations (Cadwell RC 1994) and DNA shuffling uses the combination of enzymes to first fragment the gene of interest into small pieces of DNA and then several rounds of optimized PCR to anneal and extend the DNA fragments, ultimately creating recombined portion of these genes with unique properties (Stemmer 1994). The advantages of using these procedures is that one could generate single or combinations of mutations within genes and then select for these variants using a broad range of assays to identify optimized or decreased variant binding (Stemmer 1994), activity (Chen and Arnold 1993) and stability (Zhao and Arnold 1999) *in vitro* or *in vivo*. The examination of these factors *in vitro* allows for characterization of biochemical protein properties, while *in vivo* environments provide endogenous conditions that subject protein variants to native conditions, trafficking, post-translational modifications and binding partners. Further development of surface display techniques that display protein variants on the surface of yeast cells (Boder and Wittrup 1997), ribosomes (Mattheakis et al. 1994) and phage (Smith et al. 1995) coupled to fluorescence



activated sorting enables high-throughput screening and selection of variants (Levin and Weiss 2006), followed by sequencing of variants to link phenotype to genotype.

Directed evolution is a unification of the fields of protein engineering and evolutionary biology that is used to generate and select for improved variants over a range of mutational changes and ultimately mimics the process of natural selection. Specifically, through directed evolution one subjects a gene to repeated rounds of mutagenesis to create a library of mutant variants and identify optimized protein function (Cobb et al. 2013). Once these mutants are selected for and isolated one can use them as templates for the next round of diversification, selection and repeat the process until the desired trait is attained (Cobb et al. 2013). Additionally, the advantages of using directed evolution is that one does not need to have prior knowledge of the effects of amino acid substitutions or protein structure. Moreover, directed evolution can be done in purified protein systems or model organisms to study biophysical properties of variants under endogenous conditions. Directed evolution has allowed for testing of evolutionary hypothesis and further elucidation of evolutionary mechanisms of proteins. Examples include evolvability of novel protein function (Aharoni et al. 2005; Chockalingam et al. 2005; Bloom and Arnold 2009), observation of full evolutionary trajectories (Peisajovich and Tawfik 2007; Bloom and Arnold 2009), insight into protein fitness landscapes using computational models combined with directed evolution (Aita et al. 2002), epistasis (Steinberg and Ostermeier 2016) neutral drift (Peisajovich and Tawfik 2007; Bloom and Arnold 2009), promiscuity and divergence of protein function (Bloom and Arnold 2009; Khersonsky and Tawfik 2010).

While directed evolution approaches have provided valuable ways to observe and study evolution at the protein level there are several caveats that one must take into consideration when interpreting results. For example, enhanced or decreased protein function or activity is not directly correlated to organismal fitness. Additionally, directed evolution subjects a gene to specific and isolated selective pressures and screening methods, which does not mimic what happens in the natural world. In nature biological entities evolve under diverse biological selection pressures and mutations that may give rise to enhanced or defective proteins in the lab may not have those same effects in natural settings and vice versa (Bloom and Arnold 2009). Furthermore, because one screens for optimized function to select for specific mutants one fails to illuminate impacts from many other mutations which affect protein stability, thermodynamics and function.

### **Nucleic Acid Sequencing Technology**

The purification of bacteriophages with DNA genomes and the incorporation of radioactive nucleotides by DNA polymerase one at a time was the first method that could permit the measurement of nucleotide incorporation and indirectly determine the order of nucleotides (Wu and Kaiser 1968; Wu 1970), whereas, direct identification of sequence bases and linear sequences was restricted to short DNA stretches because the only techniques available at the time were analytical chemistry and fractionation. It was not until 1977 that the breakthrough in direct DNA sequencing occurred as a result of the development of Fred Sanger's 'chain-termination' or dideoxy technique, otherwise known as Sanger sequencing (Sanger et al. 1977). Briefly, in this procedure

radiolabeled chemical analogs of deoxynucleotides (dNTPs), dideoxynucleotides (ddNTPs) are mixed into an *in vitro* DNA synthesis reaction at a fraction of the concentration of standard dNTPs, resulting in the random incorporation of ddNTPs by DNA polymerase and the production of terminated DNA fragments of varying sizes. By using a combination of heat denaturation, gel separation and autoradiography of DNA fragments one can measure the DNA fragment size and sequentially identify the nucleotide at each position, thus resulting in the ability to solve the complete linear DNA sequence (Sanger et al. 1977). Because of the accuracy, robustness and ease of use due to improvements in detection via use of fluorescently labeled ddNTPs, capillary based electrophoresis and sequencing automation, Sanger sequencing has become and is currently the most popular method for sequencing DNA of < 1,000 base pairs, and for molecular cloning. However, because Sanger sequencing provides accuracy for sequences below 1,000 base pairs one cannot rely on this technique for accurate sequencing of large and or complex genomes. To improve upon this, techniques such as shotgun sequencing, which clones overlapping DNA fragments for automated separation and sequencing followed by assembly into one contiguous sequence or 'contig" were developed to analyze longer fragments (Anderson 1981). The combined use of the shot-gun approach with automated DNA sequence detection and analysis was instrumental in the simultaneous sequencing of hundreds of samples and provided the first complete human genome sequence (Venter et al. 2001). Even though first-generation sequencing methods aided in the sequencing of DNA sequences, they were still suboptimal, time consuming and expensive.

## Next Generation Sequencing Technology

The ability to directly sequence larger and more complex genomes in a high throughput systematic manner became possible due to the production of second, or 'next' generation (Next Gen) automated sequencing technology. Briefly, this technology relies on the preparation of DNA libraries via fragmentation, purification and amplification of DNA fragments (Heather and Chain 2016). The amplified products are then isolated either through attachment to small beads or surfaces (Heather and Chain 2016). One can then resolve the sequence of each of the DNA fragments by using the inert functions of DNA such as, base incorporation during DNA synthesis (Illumina, Helicos, and SMRT) (Liu et al. 2012; Quail et al. 2012) pyrosequencing or hydrogen ion exchange during base incorporation (454 and Ion Torrent) (Liu et al. 2012; Quail et al. 2012) and base pair hybridization and ligation (Solid and NanoBall) (Heather and Chain 2016). Together these methods allow for a parallelized work flow which promotes the detection of hundreds of thousands (SMRT) to millions (Illumina Hi Seq) of DNA fragments in a mixed pool in real time. Most importantly, any alterations in sequence can further be identified in a single reaction via computational alignment to reference sequences. As a result, Next Gen sequencing has allowed for the sequencing of over hundreds of different eukaryotic organisms, which has enabled the study of the differences between organisms of different species and determine how they have evolved. In addition, Next Gen sequencing can sequence individuals within populations, which has prompted the use of this technology in combination with microbial experimental evolution to observe evolution in real time (Brockhurst et al. 2011) More

recently, high throughput sequencing of over 100,000 individual human genomes has provided insight into human genetic variation (Abecasis 2012), disease risk and enabling personalized medicine (Biesecker 2010; Ng et al. 2010; Biesecker and Peay 2013). Furthermore, Next Gen sequencing has revolutionized the way in which one could study the interactions among sequence-structure-phenotype and fitness.

### **Mutational Scanning Approaches to Reveal Protein Fitness Landscapes**

The fitness landscape or adaptive landscapes is a fundamental concept in Evolutionary Biology that was first introduced by Wright in 1932 to visualize the relationship between genotype and reproductive success or fitness for all combinations of mutant alleles in a topographical space, where height is a representation of fitness (Wright 1932). Wright's fitness landscape provided mechanistic insight into how groups of individuals within a population can continuously find its way from lower (deleterious) to higher (adaptive) fitness when subjected to the effects of mutation rate, environmental changes, strength of selection and demography. Since Wright, others have developed renditions of Wright's fitness landscape model to gain further insight into adaptive evolutionary processes (Kaufman and Levin 1987; Poelwijk et al. 2007; Kondrashov and Kondrashov 2015). Furthermore, the use of large-scale data sets such as, sequence isolates from patients and *in vitro* viral replication measurements combined with theoretical models and quantitative predictions have highlighted key properties of viral fitness landscapes (Kouyos et al. 2012; Barton et al. 2015). However, these approaches are limited because of the constrained throughput of fitness

measurements for variants and the vast sequence space that is incompletely sampled. As a result, these methods provide glimpses into partial protein fitness landscapes.

Mutational scanning is an approach that increases mutation screening throughput by using a combination of creating mutant libraries of genes, selection and Next Gen sequencing to monitor and measure the changes in mutant frequencies in bulk competition before and after selection. By measuring these changes one can then estimate the biochemical and fitness effects of all mutations across genes in genetically trackable systems including, bacteria, yeast and mammalian cells. Most importantly, one can directly link the genotype to phenotype for all possible or specific mutations and obtain empirical fitness landscapes. Using these fitness landscapes one can gain a clearer understanding of molecular mechanisms.

One of the earliest mutational scanning approaches was developed by Weiss et al., termed Alanine Scanning (Weiss et al. 2000). In this approach, Weiss used oligonucleotide-based site-directed mutagenesis to create combinatorial libraries of alanine and wild-type substitutions in 19 residues of the high-affinity binding site of human growth hormone (HGH) that were incorporated into an *E. coli* host. Phage display, shotgun library sorting and binding assays between HGH and human growth hormone binding partner (HGHbp) were subsequently used to select for variants that were sequenced to identify the nucleotide composition of variants (Weiss et al. 2000). By focusing on the distribution of alanine or wild type in each scanned position they were able to identify wild-type amino acid positions that were highly conserved and enriched for in positions with energetically favorable binding contacts between HGHbp and HGH (Weiss et al. 2000). In comparison, wild-type sidechains that did not

contribute to the binding energy of HGHbp-HGH were not enriched. While alanine mutational scanning approaches like the one exemplified can provide insight into the effects of the functional and structural properties of every amino acid in a protein it fails to highlight what occurs naturally. Specifically, in nature nucleotide mutations occur randomly throughout a gene, resulting in amino acid changes that are synonymous or non-synonymous. Additionally, one is unable to analyze the combined mutational effects from other amino acid side chains within the protein and thus fails to provide information on the position specific biophysical requirements such as, polarity and electrostatic effects. As a result, alanine scanning fails to illuminate all mutant effects, which are important for obtaining complete fitness landscapes and in construing the relationship among sequence, structure and function.

To try to circumvent the limitations of site directed or specific amino acid substitution methods many have turned to computational approaches like Condel (González-Pérez and López-Bigas 2011), GERP (Cooper et al. 2010) and SIFT (Kircher et al. 2014) to predict the effects of mutations on structure and function. These methods use statistical analysis of conservation and diversification of natural sequences and biophysical constraints of each residue to predict mutant effects. However, these approaches can only provide accurate predictions for an average of around half of mutations (Fowler and Fields 2014). As a result, these approaches fail to accurately predict all mutant effects on protein structure and function and provide a partial fitness landscape.

Recently developed deep mutational scanning (DMS) approaches that combine saturation mutagenesis techniques, bulk competition, selection of mutations based on

affinity-based isolation and deep sequencing of variants before and after selection promotes the study of the effects of all possible mutants in a gene on protein structure, stability and function *in vitro*. Specifically, chemical DNA synthesis (Fowler and Fields 2014), error prone PCR (Wu et al. 2013) or randomized cassette ligation (Hietpas et al. 2011) is used to generate most single amino acid substitutions within a gene. These variants are then pooled during bulk competition to promote analyses and selection of mutants under similar selection pressures. For example, pioneering work in the development of deep mutational scanning was done by Fowler et al., who used chemical gene synthesis with engineered degeneracy throughout most codons to create a large and complex mutant library of ~600,000 variants including, single nucleotide substitutions and a fraction of double and triple nucleotide substitutions within a human WW domain (Fowler et al. 2010). Each variant was expressed using phage display and selected for by bulk selection of mutant binding affinity to peptide ligand. Deep sequencing was used before and after selection to monitor the performance of each variant and determine the frequency change of mutants, which was directed associated with binding affinity (Fowler et al. 2010). Through this approach Fowler et al. was able to define a high-resolution protein fitness landscape map of each mutational preference across the human WW domain, ultimately shedding light on the features of each position and the relationship among sequence, protein structure and function. Furthermore, this approach can be used *in vitro* and *in vivo* to determine protein fitness landscapes for a large number and complexity of mutations within a specific gene and reveal biophysical, structural and functional mechanisms that allow for a further understanding of molecular adaptations. For example, modifications of this approach



have been used in the design of novel influenza inhibitors (Whitehead et al. 2012), determination of protein half-lives and stability for the *S. cerevisiae* degradation signal, Deg1 and the human WW domain (Araya et al. 2012; Kim et al. 2013), uncovering activity enhancing mutations in murine E3 ubiquitin ligase (Starita et al. 2013), interdependency between mutations in the same gene for the WW domain (Araya et al. 2012) and epistatic interaction networks in the green fluorescent protein of *Aequorea victoria* and the RRM domain of the *S. cerevisiae* poly A binding protein (Melamed et al. 2013; Sarkisyan et al. 2016). These studies have allowed for the high throughput study of the impact of mutations on protein stability and function (enzymatic or binding activity) *in vitro* and ultimately reveal experimentally derived protein fitness landscapes.

To quantify accurately and systematically the effects of all mutations within a gene on organismal fitness *in vivo*, Hietpas et al., developed a similar yet distinct deep mutational scanning approach, termed “extremely methodical and parallel investigation of randomized individual codons” (EMPIRIC) fitness. In this approach Hietpas et al., 2012 used randomized cassette ligation to introduce all codon substitutions including, single, double and triple mutations within a nine amino acid region of the yeast Heat shock protein (*hsp82*) gene, which is the yeast version of Hsp90. Hietpas incorporated these cassettes into a plasmid-based system (Hietpas et al. 2012). These plasmid libraries of Hsp90 mutations were transformed into yeast, followed by bulk competition to select for enriched and depleted variants at various timepoints throughout competition (Hietpas et al. 2012). Next-gen deep sequencing was then used to monitor the relative abundance of each mutant in bulk culture over time. Analysis of the change in frequency of Hsp90 mutants over time versus wild type Hsp90 allows for

an estimation of the impact of Hsp90 mutations on yeast growth rate, otherwise known as experimental fitness (Hietpas et al. 2012). Using this approach Hietpas was able to directly map the sequence-fitness relationship for this nine amino acid region of yeast Hsp90 and provide further insight into evolutionary mechanisms and adaptation. For example, Hietpas discovered that the distribution of mutant fitness effects for single mutations are bimodal under standard conditions with most mutations demonstrating either mostly deleterious or neutral effects on fitness, providing experimental evidence in support of a nearly neutral model of evolution (Hietpas et al. 2011). In another EMPIRIC study they analyzed the effects of shifted environmental conditions (e.g., elevated salinity and increased temperature) on the Hsp90 fitness landscape and identified environment specific adaptive Hsp90 mutations that demonstrate costs of adaptation between mutations and are consistent with Fisher's Geometric Model of adaptation (Hietpas et al. 2013). Modifications of the EMPIRIC approach have been applied to study the effects of varied expression levels on the Hsp90 fitness landscape (Jiang et al. 2013), to uncover an intragenic epistatic landscape in combined mutations of yeast Hsp90 (Bank et al. 2015), analysis of the effects of ubiquitin mutations on yeast growth rate (Roscoe et al. 2013) and E1 activation (Roscoe and Bolon 2014) and determining the yeast ubiquitin fitness landscape under diverse environmental conditions (Mavor et al. 2016). EMPIRIC has also been used extensively to study viral evolution and drug resistance including, determining the fitness effects of all point mutations and synonymous substitutions in a region of the Influenza A viral protein, hemagglutinin (Canale et al. 2018), quantifying the effects of all single-nucleotide substitutions in important regions of the Influenza A viral protein, neuraminidase (Jiang

et al. 2016) and determining the fitness landscape of single amino acid substitutions in HIV protease (Boucher et al. 2019) and the HIV-1 envelope CD4 binding loop (Duenas-Decamp et al. 2016). Most importantly all together EMPIRIC has allowed for an in-depth study of fitness landscapes based on the impact of mutations on *in vivo* organismal growth rate, which are mediated by changes in protein properties. These studies have shed light on important concepts in molecular evolution and adaptation.

More recent studies have used DMS to investigate the relationship between gene sequence and function under environmental-dependent selection pressures. For example, many studies have used DMS to investigate drug or antibody resistance mutations that provide a growth advantage under these environments (Firnberg et al. 2014; Stiffler et al. 2015; Doud and Bloom 2016; Dingens et al. 2018). While Dandage et al., used DMS to examine the effect of environmental perturbations on mutations within the Gentamycin resistance gene in *E. coli* to determine how different biophysical parameters of mutations constrain molecular function in different environments (Dandage et al. 2017). Studies of this nature have provided observations that imply how changes in protein structure as a result of mutations can promote or constrain adaptation to different environments.

Even though DMS approaches have been used extensively to address protein-to-sequence- function, molecular evolution and adaptation there are multiple limitations that one needs to consider when interpreting results. For example, experimental observations do not coincide with natural sequence conservation patterns (Hietpas et al. 2012; Wu et al. 2015; Flynn et al. 2020). These differences could be due to various biological and technological factors including, effective population size (Ohta 1973),

differences in the resolution between experimental evolution studies and natural evolution (Reznick and Ghalambor 2005) and differences in natural versus experimental selection pressures (Mustonen and Lässig 2009). Furthermore, because DMS relies on deep sequencing there are two technological factors that can impact the scale and accuracy of sequencing results, which are read depth and read errors (Laehnemann et al. 2016). Both factors can result in sampling noise within sequencing results that prevent accurate interpretation of results. To overcome these challenges, sequencing samples to sufficient read depth, using paired-end sequencing and the use of barcoded indices have improved the quality of sequencing results (Hiatt et al. 2010; Starita et al. 2013). Applying this technology to the EMPIRIC method could promote accurate measurement of the impact of several environmental stresses on the distribution of mutant fitness effects within a large stretch of a functional region of yeast Hsp90 and the entire Hsp90 sequence and can inform how environment impacts Hsp90 mutant fitness effects, evolutionary constraints, adaptive potential and evolution of Hsp90.

## **Fitness**

In population genetic models Fitness ( $W$ ) (Haldane 1927) is a quantitative measurement of an individual's competitive advantage/disadvantage that is used to describe an individual's reproductive success within an environment (Haldane 1927). As a result, fitness can represent a quantitative measurement of natural selection. Fitness can be measured in two ways either as absolute or relative. Absolute fitness is based on direct measurement of the number of individuals possessing a specific genotype in a population before and after selection (Kimura 1970), whereas, relative fitness is based

on the measurement of an individual's growth / reproductive rate compared to the population's average growth/reproductive rate in a single generation (Kimura 1970). Relative fitness is used throughout this dissertation when analyzing and discussing fitness effects of yeast Hsp90 mutations and their evolutionary implications.

Genetic mutations are a primary mechanism that can impact an individual's phenotype/fitness, therefore the interactions between mutations and fitness can provide insight into adaptive evolutionary potential. Additionally, because selection pressures can have direct evolutionary consequences on gene sequence the relationship between these two factors can provide a means to connect genotype and fitness. As a result, many have investigated the fitness effects of mutations using diverse approaches including, population genetics models combined with polymorphism data (McDonald and Kreitman 1991; Boyko et al. 2008; Schneider et al. 2011; McDonald 2019), which makes fitness inferences based on sequence analysis of natural evolving populations. Studies of this nature promote an understanding of recent selection in sequenced organisms. While others have used microbial experimental fitness competitions to directly measure the selective growth advantage/disadvantage of specific mutants under specific selective pressures (Kassen and Bataillon 2006; Rozen et al. 2007; McDonald et al. 2012; McDonald 2019). While, both approaches have provided distinct ways to obtain fitness inferences for mutants, there are caveats to using both approaches. For example, population models combined with polymorphism data is limited to recent selection within naturally evolved populations, thus one cannot determine the pool of starting mutations before selection. Furthermore, mutations that cause a selectable effect on fitness can be difficult to ascertain because of hitchhiking

mutations at linked genetic loci that may also have an effect on fitness (Smith and Haigh 1974). In terms of experimental fitness measurements these studies rely on isolation of specific mutants and following their growth for multiple generations under specific environments. Therefore, these studies are limited to microbial model systems because they have short generation times and their genomes are easily manipulated.

Furthermore, because experimental fitness measurements rely on isolation of specific mutants mutational sampling is limited. To overcome challenges in measuring the fitness of mutants Hietpas et al., developed the EMPIRIC fitness approach to accurately and systematically measure the competitive growth advantage or disadvantage of single substitutions in bulk competition in a high-throughput manner (Hietpas et al. 2011). In this approach wild-type sequences are included in the mutant library to enable the measurement of the change in ratio of mutant to wild-type sequence reads over time during bulk competition (Hietpas et al. 2011). The relative fitness of each mutant is calculated as selection coefficients ( $s$ ) and represents the difference in fitness between mutant and wild-type over time (Hietpas et al. 2011). Because yeast fitness is proportional to the inverse of the doubling time the fitness ( $s$ ) of wild type yeast is equal to 1, therefore a selection coefficient of zero indicates no change in mutant to wild-type ratio indicating that a mutant is wild-type like, whereas, a negative selection coefficient indicates that a mutant is less fit than wild-type, and a positive selection coefficient means that a mutant is more fit than wild-type (Hietpas et al. 2011). These mutant fitness calculation methods are used in this dissertation to distinguish mutant fitness effects. Together EMPIRIC has allowed for the accurate measurement of the fitness of relatively large numbers of point mutations in various proteins under diverse selection

pressures (Hietpas et al. 2013; Roscoe and Bolon 2014; Jiang et al. 2016; Mavor et al. 2016; Canale et al. 2018) and revealed large scale protein fitness landscape maps.

### **Distribution of Mutant Fitness Effects**

Mutation is the source of all genetic variation within an individual's genome and is the primary method by which natural selection acts to promote evolutionary change within individuals and adaptation to new environments. However, most mutations cause negative effects on an organism's growth and reproduction, termed 'fitness' and are removed by purifying selection (Ohta 1973). Additionally, many mutations have neutral effects on fitness that are affected by stochastic processes like genetic drift (Ohta 1973), whereas, beneficial mutations that provide a fitness advantage are rare but are subsequently selected for because they contribute to evolutionary change (Ohta 1973). As a result, quantifying the distribution of fitness effects (DFE) of new mutations allows one to identify the proportions of beneficial, deleterious or neutral mutations within a gene and gain an understanding of how the proportions of these mutations shape evolution. For example, determining the DFE's of new mutations has provided insight into evolutionary mechanisms, such as the maintenance of molecular genetic variation (Charlesworth et al. 1995), the 'molecular clock' (Ohta 1992) and the evolution of sex and recombination (Peck et al. 1997). Moreover, DFEs allows one to predict evolutionary dynamics, like estimating the frequency of beneficial mutations and revealing the adaptive potential of these mutations within a new environment (Sniegowski and Gerrish 2010; Bataillon and Bailey 2014).

The theoretical population geneticist, Motoo Kimura was one of the first to calculate the fixation of the proportions of neutral, beneficial and deleterious mutations within a gene (Kimura 1968). Through his calculations and comparisons with molecular evolution data Kimura introduced an important molecular evolutionary concept termed the “neutral theory of molecular evolution”. The neutral theory of molecular evolution holds that random genetic drift of selectively neutral mutant alleles results in molecular evolutionary changes and contributes to variation within and between species (Kimura 1983). Furthermore, most mutations are deleterious and quickly purged as a result of natural selection, and thus do not contribute to molecular evolution and variation (Kimura 1983). Together the Neutral Theory of Molecular Evolution is an important concept that is compatible with Darwin’s Theory of Natural Selection shaping phenotypic evolution. Furthermore, it has served as the basis for other important theories including the “nearly-neutral theory of molecular evolution” by Ohta and Kimura, which is a modification of the “neutral theory of molecular evolution” and takes into consideration that not all mutations are severely deleterious (Ohta 1992). Furthermore, it demonstrates that the population dynamics of these mutations are not much different than neutral mutations indicating the potential role of these mutations in evolution.

A variety of experimental approaches have been used to determine the DFEs in genetically trackable model systems including: theoretical models (Burch et al. 2007), laboratory experimental evolution, whereby microbial or viral populations are evolved in the lab and spontaneous mutations are tracked (Lenski et al. 1991; Wloch et al. 2001; Sousa et al. 2012) and quantitative methods that provide statistical inference from



natural polymorphism and/or divergence data (Nielsen and Yang 2003; Piganeau and Eyre-Walker 2003; Keightley and Eyre-Walker 2010). A more direct approach to investigate the DFE of new mutations is to use directed evolution approaches such as, site directed mutagenesis to induce new mutations and measure the fitness effects of each variant. For example, this method has been used to determine the DFEs of new mutations and the mutation rate in yeast (Wloch et al. 2001), DFEs of random insertions in bacteria (Elena 1998), and DFEs of single nucleotide substitutions in an RNA virus to estimate the abundance and effects of nearly neutral mutations in RNA viruses (Sanjuán et al. 2004). These approaches have provided valuable insight into DFEs of novel mutations and further predict evolutionary dynamics such as mutation rate and the abundance of nearly neutral mutations. However, these methods are constrained because accurate measurement of fitness effects are dependent upon large effects on fitness,  $> 1\%$  (Eyre-Walker and Keightley 2007). Therefore, one cannot obtain a complete inference of the DFE of new mutations within a specific genomic region. To overcome these limitations many have turned to site-directed mutagenesis approaches combined with deep sequencing to obtain complete and close to accurate DFEs for an entire gene or a region of a gene in yeast (Fowler et al. 2010; Hietpas et al. 2011) and viruses (Boucher et al. 2019). These studies have also revealed a bimodal DFE for new mutations, with one peak centered around neutral or wild-type like mutations and a second peak centered around strongly deleterious mutations, consistent with Ohta's nearly neutral model (Ohta 1992). More recently, site-directed mutagenesis combined with deep sequencing fitness scans have allowed for the study of the changes of the DFE of new mutations in response to novel environmental stresses and has uncovered

adaptive mutations in Hsp90 that are adaptive under one environmental condition but deleterious in another, otherwise known as “costs of adaptation” (Hietpas et al. 2013; Bank et al. 2014). Together these studies highlight the importance of determining the DFE of new mutations to gain new perspectives on evolutionary mechanisms and dynamics and warrant future studies of the relationship between environment and DFE of new mutations.

## The Heat Shock Protein 90

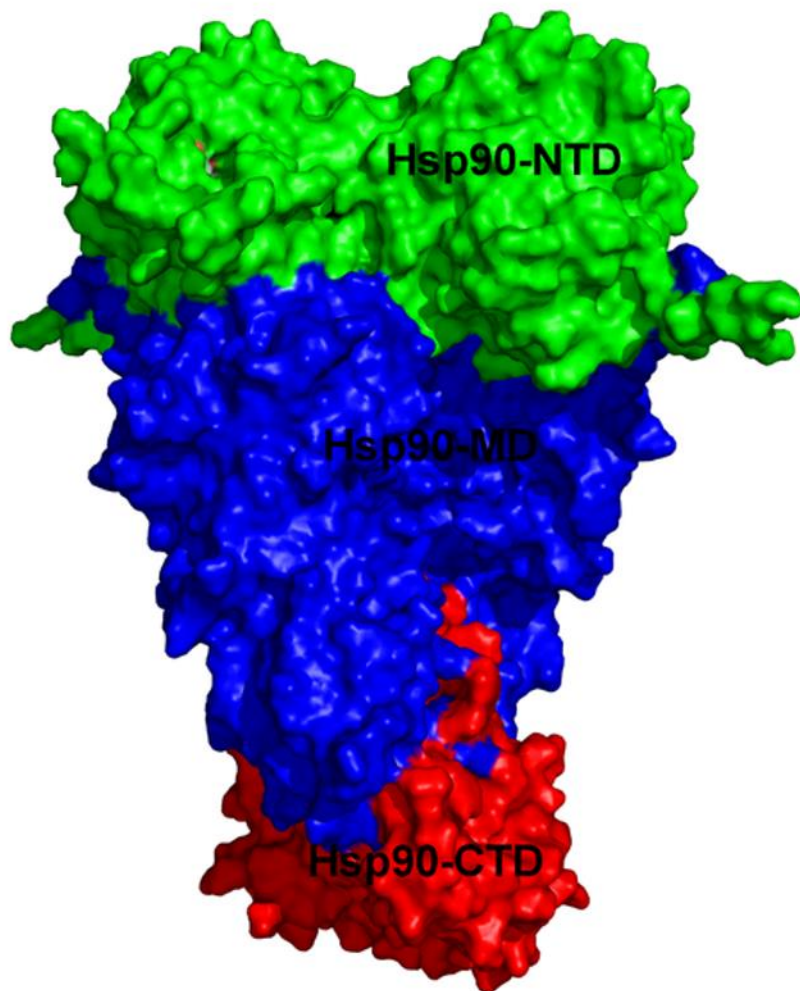
### Heat Shock Protein 90 Function and Structure

The heat shock protein, Hsp90 is a highly conserved and essential molecular chaperone that is expressed in a variety of different organisms from bacteria to mammals. Hsp90 was first discovered in *Drosophila melanogaster* larvae after exposure to high temperatures (McKenzie et al. 1975). In *S. cerevisiae*, Hsp90 is expressed in the cytosol as two closely identical isoforms (97% identity), *Hsp82* (Hsp90) and *Hsc82* (Hsc90) (Borkovich et al. 1989). Both isoforms are constitutively expressed, however Hsp90 is inducible under conditions that induce proteotoxic stress, such as increased temperature (Borkovich et al. 1989). Under normal conditions Hsp90 regulates diverse cellular processes through its association with clients. Previous large-scale proteomic, genetic and biochemical studies in yeast and mammalian systems have shown that Hsp90 interacts with an astonishing number of clients including, kinases, E3 ligases, and transcription factors (Zhao and Arnold 1999; McClellan et al. 2007; Taipale et al. 2012) (<https://www.picard.ch/downloads/Hsp90interactors.pdf>). Hsp90 association with

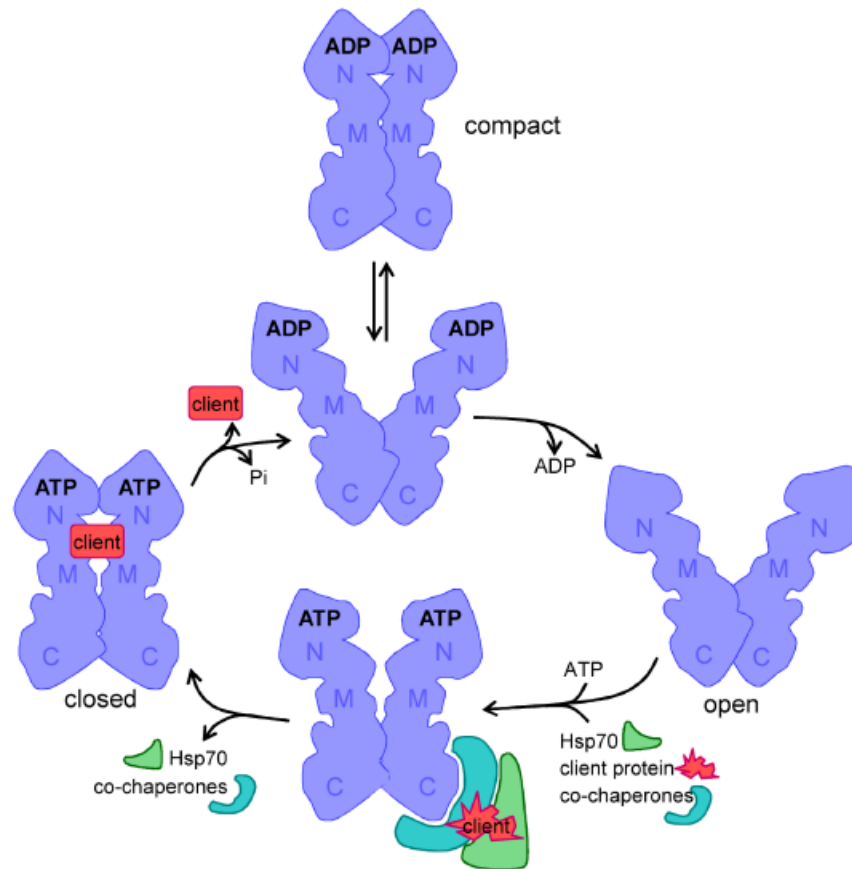
these clients is required to facilitate their refolding, stabilization and maturation into conformationally active proteins(Li 2012). The maturation of these clients in turn regulates diverse processes including cell cycle control, signal transduction, growth, metabolism and transcriptional reprogramming (McClellan et al. 2007). Because Hsp90 binding to clients facilitates the maturation of clients into active conformations that regulate most aspects of cellular biology, Hsp90 has the potential to impact cellular and organismal physiology. However, the intrinsic mechanism by which Hsp90 interacts with these clients and the biological significance of these interactions remains unclear and thus has stimulated extensive research into how this chaperone functions on many levels.

Hsp90 is a flexible homodimeric protein, which consists of three domains in each monomer, the N-terminal- ATP binding domain (N), a middle domain (M) and a C-terminal dimerization domain (C) (Figure 1.1) (Rohl et al. 2013). While the mechanism by which Hsp90 remodels clients is not completely understood, previous structural, biochemical and empirical studies of the N-terminal domain have shown that ATP binding to the N-terminal domain of Hsp90 and ATP hydrolysis facilitates large conformational rearrangements essential for Hsp90 function and maturation of most clients (Panaretou et al. 1998; Mishra et al. 2016). Furthermore, *in vivo and in vitro* functional analyses of Hsp90 have demonstrated that the ATP coupled-conformational cycle of Hsp90 along with accessory proteins (co-chaperones) are required to promote conformational changes that allow Hsp90 to bind to partially folded clients, stabilizing specific folding intermediates that in turn allows clients to attain their full function (Figure 1.2) (Leach et al. 2012) (Lotz et al. 2003; Roe et al. 2004; McLaughlin et al. 2006;

Onuoha et al. 2008). For example, previous functional analysis of Hsp90 interactions with model vertebrate clients such as, the oncogenic transforming kinase, v-Src and the Glucocorticoid Receptor (GR) in yeast cells, human cells and yeast Hsp90 knockouts complemented with human Hsp90 have shown that Hsp90 is required to prevent the degradation and mediate the activation of these clients (Picard et al. 1990; Xu and Lindquist 1993; Whitesell et al. 1994; Mishra et al. 2016). Moreover, these studies have shed light on the conservation of Hsp90 function between yeast and humans and have highlighted the value of studies that utilize yeast Hsp90 and model clients to further understand Hsp90 mechanism. Additionally, EMPIRIC studies of yeast Hsp90 have identified a panel of Hsp90 N-terminal mutations with severely compromised function (Mishra et al. 2016). These mutations were in close proximity to the ATP binding site. Therefore, the ATP coupled-chaperone cycle of Hsp90 is required for Hsp90 function in client maturation and is essential for the activity and regulation of fundamental intracellular processes that ultimately influence physiology, growth and disease.



**Figure 1.1: Yeast Hsp90 structure:** Modified Hsp90 structure from (Blacklock and Verkhivker 2013).



**Figure 1.2: Current model of Hsp90 mediated client maturation:** Hsp90 mediates client maturation via ATP binding and ATP hydrolysis. ATP binding to the N-terminal domain of Hsp90 facilitates N-terminal dimerization and large conformational changes required for Hsp90 interaction with cochaperone client complexes, client transfer and maturation. Subsequent ATP hydrolysis is essential for client release. Figure adapted and modified from (Leach et al. 2012).

The mechanism by which Hsp90 physically interacts with clients and promotes their remodeling remains unclear because of its extensive binding interface, the diverse repertoire of clients, intrinsic instability and transient client-binding interactions (Rohl et al. 2013). Previous structural studies of Hsp90 in combination with mutational analysis

have uncovered specific hydrophobic residues or hydrophobic patches extended throughout the N-terminal, middle and C-terminal domain, which contributes to the recognition and interaction with specific clients (Rohl et al. 2013). However, the middle domain of Hsp90 contains the majority of hydrophobic patches and has been found to bind to most clients indicating its primary role in Hsp90 client binding.

The M domain is the largest domain of Hsp90, consisting of amino acids 255-599 in *S. cerevisiae* (Meyer et al. 2003) and is strongly conserved (64% identity from yeast to human). While all domains have been implicated in client binding the M domain has been shown to contain multiple client binding hotspots in various organisms (Rohl et al. 2013). Biochemical studies of mutations within the middle domain of yeast Hsp90 have identified conserved residues that are important in the binding and maturation of vertebrate clients such as, Glucocorticoid Receptor and v-Src (Nathan et al. 1997; Meyer et al. 2003; Hawle et al. 2006). Consistent with Hsp90's critical role in binding and maturation of these clients, many of these Hsp90 mutations were found to also cause growth defects (Nathan and Lindquist 1995; Meyer et al. 2003). Other biochemical studies of mammalian Hsp90 identified a region within the middle domain that selectively binds to both protein complex B (Akt) and apoptosis regulating kinase 1 (Ask1) which inhibits hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>)-induced ASK1–p38 activation in endothelial cells (Zhang et al. 2005). Similarly, analysis of human Hsp90  $\beta$  have identified M domain amino acids to be involved in the binding to Akt (Sato et al. 2000). In addition to mutational, biochemical and growth analysis, structural studies have helped to elucidate the important role the middle domain plays in binding to various Hsp90 clients. For example, earlier structural studies of the middle domain of Hsp90

have detected numerous hydrophobic patches scattered throughout the middle domain, including a solvent exposed hydrophobic patch that is part of an amphipathic loop, both of which are indicative of protein-protein interaction sites (Meyer et al. 2003). Recent structural studies of human Hsp90 have further shown that the middle domain binds to many clients including Tau (Karagoz et al. 2014), the DNA binding domain of p53 (Hagn et al. 2011; Park et al. 2011) and binding of chaperone-client complexes Cdc37-Cdk4 (Verba et al. 2016; Czemerer et al. 2017). Together these studies provide evidence of the important role the middle domain of Hsp90 plays in client binding and maturation, ultimately affecting cellular and organismal homeostasis. As a result, chapter two applies mutational studies of part of the middle domain to extend our knowledge of this region and its role in evolutionary biology.

### **Hsp90 and Adaptation to Stressful Environments**

Natural environments change frequently due to the dynamic alterations of abiotic factors including, temperature, salinity, and pH, which can have widespread evolutionary consequences (Elena and Lenski 2003; Dhar et al. 2011; Arribas et al. 2014). Therefore, individuals are constantly subjected to a wide range of novel stresses that reduce viability. For example, increased temperature can cause protein unfolding and aggregation, which can disrupt function, ultimately affecting organismal fitness and survival (Richter et al. 2010). Organisms have evolved cellular mechanisms, including molecular chaperones (Hsp70, Hsp90 and Hsp100) to deal with stress and facilitate environmental resilience (Kaplan and Li 2012). Hsp90 plays an essential role in protecting cells from environmental stress and is required for eukaryotic growth during adaptation to stress. For example, modest increases in temperature cause an increase



in the concentration of unfolded proteins in the cell, which in turn can aggregate and lead to proteotoxic effects that disrupt protein and cellular homeostasis (Richter et al. 2010). This heat induced protein damage can result in the disruption of the internal organization of the cell and fragmentation of organelles, leading to cell death (Sangster et al. 2004). As a result, organisms have evolved cellular adaptive mechanisms, such as the induction of the molecular chaperones (Hsp70, Hsp90 and Hsp100) to deal with proteotoxic stress and facilitate resistance via refolding of unfolded proteins (Kaplan and Li 2012). Genetic and biochemical analysis of these heat shock proteins in *S. cerevisiae* have shown that under normal growth conditions *hsp82* (Hsp90) is constitutively expressed, constituting 1-2% total cytosolic protein, and is further upregulated 20-30-fold during elevated temperatures (Borkovich et al. 1989). The upregulation of Hsp90 is required to facilitate the renaturation and recovery of heat-induced denatured protein clients, maintaining protein homeostasis and cellular function (Rutherford et al. 2007). Specifically, this is done via the dissociation of the heat shock transcription factor, Hsf1 from Hsp90 in the cytoplasm (Prodromou 2016). Hsf1 is then able to homotrimerize and translocate to the nucleus to induce the expression of Hsp90 via its interaction with three heat shock element motifs in the Hsp90 promoter (Prodromou 2016). When conditions are normal, Hsf1 dissociates from the Hsp90 promoter and translocates out of the nucleus into the cytoplasm where it can reassociate with Hsp90 and become negatively regulated (Prodromou 2016).

Hsp90 is also directly involved in regulating the yeast short-term adaptive gene expression response to other environmental stressors. Yeast short-term adaptation to novel environmental challenges involves rapid and transient changes in signaling

pathways that dramatically alter the transcription of specific subsets of genes that coordinate growth to stress responses (Nollen and Morimoto 2002). For example, previous gene expression studies in yeast grown under diverse environmental stressors including nitrogen depletion, hyper-osmotic stress (increased salinity or sorbitol) and oxidative stress (diamide and H<sub>2</sub>O<sub>2</sub>) requires global transcriptional changes that results in the upregulation of specific subsets of genes required for yeast growth (Duch et al. 2012) (Gasch et al. 2000) (Posas et al. 2000). For example, all these conditions except for osmotic stress result in the increased expression of genes encoding heat shock proteins, including Hsp90 (Gasch et al. 2000), whereas, osmotic stress results in the increased expression of genes involved in glucose production (Gasch et al. 2000). In addition, adaptation to these environmental challenges requires the increased expression of ~300 environmental stress responsive (ESR) genes involved in a wide array of cellular processes, including cell growth, signal transduction and metabolism (Gasch et al. 2000) (Posas et al. 2000). The induction of many of these environment specific and general ESR genes are regulated by Hsp90 clients including, kinases and transcription factors. For example, the induction of osmostress specific genes is regulated by the activation of the high osmolarity glycerol (HOG) pathway MAPK, Hog1, (Hawle et al. 2007; Saito and Posas 2012), whereas, the increased expression of environmental stress responsive genes involved in metabolic pathways is regulated by the activation of glycogen synthase kinase 3 (GSK3) (Hirata et al. 2003; Stankiewicz and Mayer 2012). Hsp90 has been found to be directly involved in the activation or maturation of both of these kinases and the transcription factor, Hsf1 is involved in promoting the induction of heat shock proteins (Yang et al. 2006; Stankiewicz and

Mayer 2012). Therefore, in the following chapters we investigate the effects of Hsp90 mutations on yeast short-term adaptation to conditions directly associated with increased expression of Hsp90 including, thermal stress (37°C), oxidative stress (H<sub>2</sub>O<sub>2</sub> or diamide), ethanol stress and nitrogen depletion (Gasch et al. 2000) or conditions associated with Hsp90 client binding-maturation function in stress response pathways such as, the HOG pathway induced by hyperosmotic stress (NaCl or sorbitol) (Gasch et al. 2000; Yang et al. 2006).

### **Hsp90 Facilitates the Rapid Evolution of Novel Traits**

The ability for organisms to rapidly adapt to unpredictable environments is dependent upon the presence of pre-existing genetic variation and cellular mechanisms that couple unexpected environmental changes to the evolution of adaptive traits (Sangster et al. 2004). The environmental responsiveness of Hsp90 combined with the many fundamental signal transduction pathways it mediates are factors that provide Hsp90 with unique opportunities to contribute to evolutionary change (Lindquist 2009). Therefore, much time and effort has been spent understanding the role this chaperone plays in promoting evolutionary change. Previous studies in yeast, flies, and plants have demonstrated that Hsp90 function can directly contribute to the rapid evolution of adaptive traits (Jarosz and Lindquist 2010). Under optimal conditions Hsp90 acts as a capacitor of phenotypic variation, suppressing the phenotypic consequences of mutations in signal transduction pathways by correctly folding mutant clients (Sangster et al. 2004; Jarosz and Lindquist 2010). However, during conditions of environmental stress Hsp90 buffering capacity becomes taxed as a result of the increase in the concentration of stress-induced unfolded and damaged proteins, which compete for

Hsp90 chaperone function (Sangster et al. 2004; Taipale et al. 2010). Therefore, Hsp90s ability to fold and stabilize its repertoire of clients becomes compromised, allowing the release of cryptic genetic variants and the evolution of new traits (Rutherford 2003) as demonstrated in variation in fly (Rutherford and Lindquist 1998) and plant morphology (Queitsch et al. 2002), fungal drug resistance (Cowen and Lindquist 2005) and cave fish vision loss (Rohner et al. 2013).

Recent studies using systematic mutagenesis, bulk competition and deep sequencing fitness scans have begun to address how mutations to a highly conserved client binding site of Hsp90 can impact yeast evolutionary adaptation to diverse environmental stresses (Hietpas et al. 2013). This study shows that multiple mutations in a client-binding site of yeast Hsp90 can provide a growth advantage for yeast under elevated salinity conditions (Hietpas et al. 2013). Hietpas et al., identified 13 mutations that reproducibly confer a 7-10% increase in growth rate under elevated salinity conditions in comparison to wild type Hsp90, indicating that this region may play an important role in adaptation to this condition. In addition, the identified costs of adaptation for these beneficial mutations, whereby these mutations demonstrate deleterious effects under standard and high temperature conditions (Hietpas et al. 2013).

### **Thesis Scope**

Natural environments are constantly changing, which can result in the manifestation of environmental challenges that place selective pressures upon an individual's fitness at many biological levels, including the sequence level. As a result, the interactions between genes and environments is theorized to play a significant role

in shaping molecular evolution and adaptation (Fisher 1931; Wright 1932). However, little is experimentally known about the impact of new environmental challenges on shaping the selection pressure on a protein sequence and their evolutionary consequences. Quantifying the impact of environment on the DFE of new mutations can determine the proportions of neutral, deleterious and beneficial mutations under specific environments and thus provide a glimpse into how environmental change impacts selection of protein sequences and their evolutionary consequences including, how does environment impact the shape of the DFE, adaptive potential, and costs of adaptation. However, very few studies have accurately quantified the impact of environment on the DFE of mutations, adaptive potential and costs (Hietpas et al. 2013; Bank et al. 2014) due to difficulties in measuring the fitness effects of the same mutations across environments. As a result, the foundation of the work presented in this dissertation focuses on applying a high throughput systematic mutagenesis approach to accurately measure the impact of diverse environmental stresses on the DFE of new mutations in yeast Hsp90 to address questions as to how environment impacts the shape of the DFE of new Hsp90 mutations, adaptive potential of Hsp90, costs of adaptation and selection of Hsp90 gene sequences in alternate environments. Yeast Hsp90 was chosen for this work because yeast Hsp90 is important for the short-term response to new and periodic environmental challenges and facilitates genetic adaptation to new stress conditions. Furthermore, previous studies have identified that mutations to Hsp90 can promote adaptation to specific environments.

In chapter II I describe how I applied an optimized version of the EMPIRIC approach to create single point mutation libraries within a highly conserved 120 amino

acid sequence region of the middle domain of yeast Hsp90, implicated in client-binding. Using EMPIRIC, I measured the growth effects of these mutants under conditions known to affect yeast growth rate and Hsp90 expression including, standard, oxidative ( $H_2O_2$  or Diamide) and hyperosmotic (NaCl, Sorbitol) stress conditions. We quantified the impact of environmental changes on the overall shape of the DFE, compared the experimental DFE to natural amino acid variants within 261 extant eukaryotes, identified regions that showed the largest proportions of beneficial or deleterious mutations, respectively, quantified hotspots of costs of adaptation, and compared the identified regions with known client binding or other structurally important sites to connect the phenotypic and fitness effects of mutation. I present findings that complement previous DFE studies (Hietpas et al. 2012; Hietpas et al. 2013; Bank et al. 2014; Boucher et al. 2016) and predictions made by Ohta (Ohta 1992) and Kimura (Kimura 1983) regarding the distribution of fitness effects in the context of the nearly neutral and neutral model of evolution. I also present findings on the adaptive potential of this region across environments and their costs of adaptation under alternate environments. I discuss the adaptive potential of this region in the light of Fisher's Geometric model. Together I present results that provide information regarding the role and adaptive potential of a large region of the middle domain of Hsp90 that supports and extends known structural properties of this region across diverse environments.

In Chapter III I used the optimized EMPIRIC method to quantify the growth effects of all Hsp90 point mutations within the entire 709 amino acid sequence of Hsp90 under conditions known to affect yeast growth and Hsp90 expression including, standard conditions and six stress conditions (diamide, ethanol, nitrogen deprivation,

hyperosmotic shock (NaCl) and temperature shock, 37°C). Low Hsp90 expression was used as a sensitive readout of Hsp90 function (Jiang et al. 2013). Using this approach, we quantified fitness maps for each condition, measured each positions sensitivity to mutation under each environment, identified environment specific beneficial and deleterious mutations, examined environmentally responsive mutations for structural and physical patterns and compared with known client binding or other structurally important sites, examined the adaptive potential of the full DFE under all environments using Fisher's Geometric Model and finally compared our experimental fitness effects to the historical record of hundreds of Hsp90 substitutions accrued during its billion years of evolution in eukaryotes. We present comprehensive fitness maps for the entire Hsp90 protein sequence at low expression levels for six environments. We present findings that indicate that environment can have a large impact on the evolution of Hsp90. Specifically, we present results that demonstrate that specific environments have distinct effects on the selection of beneficial and costly Hsp90 mutations. These results coincide with previously identified high costs of adaptation for beneficial mutations in a smaller region of Hsp90 (Hietpas et al. 2013), and indicates that naturally fluctuating environments may reduce or eliminate positive selection of Hsp90. We present distinct structural trends for mutations that provide environment-dependent benefits and costs, which reveals how mutations may impact biochemical function and evolutionary mechanism. All together we present findings that indicate that environment can impact Hsp90 evolution but that this type of evolutionary mechanism is rare and suggests that natural selection on Hsp90 sequence is governed by strong purifying selection integrated over multiple stressful conditions.

A summary and discussion of my results are presented in Chapter IV.



## Chapter II: The Adaptive Potential of the Middle Domain of Yeast Hsp90

This work has been previously published as *Pamela A. Cote-Hammarlof*<sup>1\*</sup>, *Inês Fragata*<sup>2\*§</sup>, *Julia Flynn*<sup>1</sup>, *David Mavor*<sup>1</sup>, *Konstantin B. Zeldovich*<sup>1</sup>, *Claudia Bank*<sup>2,3#</sup>, and *Daniel N.A. Bolon*<sup>1#</sup>. **The adaptive potential of the middle domain of yeast Hsp90.** (\*equal contribution; # co-corresponding author) [published online ahead of print, 2020 Sep 1]. *Mol Biol Evol.* 2020;msaa211.doi:10.1093/molbev/msaa211

This was a collaborative effort. I, Pamela Cote-Hammarlof generated the concept and experimental design for this study. I performed growth rate analysis of WT yeast Hsp90 under conditions used in this experiment. I generated the mutant libraries and barcoded them. I isolated DNA from mutant libraries for sequencing and to map mutations to barcodes. I transformed yeast with mutant libraries and carried out the growth competitions with assistance from Dr. Konstantin Zeldovich and Dr. Julia Flynn. I isolated DNA from samples and prepared them for deep sequencing. I sequenced the samples and ran the initial analysis to obtain the number of barcodes associated with each mutant for each condition. Dr. Ines Fragata and Dr. Claudia Bank performed log linear regression analysis on the data I generated. Dr. Ines Fragata and Dr. Claudia Bank computed the costs of adaptation associated with adaptive mutations across environments. I, Dr. Ines Fragata, Dr. David Mavor, Dr. Claudia Bank and Dr. Dan Bolon equally contributed to analyzing the data. Dr. Konstantin Zeldovich created the turbidostat to automatically run our growth competition experiments under the conditions used in this study. I, Dr. Ines Fragata, Dr. Claudia Bank and Dr. Dan Bolon equally contributed to writing, reviewing and editing the manuscript.

**Abstract**

The distribution of fitness effects (DFE) of new mutations across different environments quantifies the potential for adaptation in a given environment and its cost in others. So far, results regarding the cost of adaptation across environments have been mixed, and most studies have sampled random mutations across different genes. Here, we quantify systematically how costs of adaptation vary along a large stretch of protein sequence by studying the DFEs of the same  $\approx 2300$  amino-acid changing mutations obtained from deep mutational scanning of 119 amino acids in the middle domain of the heat-shock protein Hsp90 in five environments. This region is known to be important for client binding, stabilization of the Hsp90 dimer, stabilization of the N-terminal-Middle and Middle-C-terminal interdomains, and regulation of ATPase-chaperone activity. Interestingly, we find that fitness correlates well across diverse stressful environments, with the exception of one environment, diamide. Consistent with this result, we find little cost of adaptation; on average only one in seven beneficial mutations is deleterious in another environment. We identify a hotspot of beneficial mutations in a region of the protein that is located within an allosteric center. The identified protein regions that are enriched in beneficial, deleterious, and costly mutations coincide with residues that are involved in the stabilization of Hsp90 interdomains and stabilization of client binding interfaces, or residues that are involved in ATPase chaperone activity of Hsp90. Thus, our study yields information regarding the role and adaptive potential of a protein sequence that complements and extends known structural information.

## Introduction

The distribution of fitness effects (DFE) determines the proportions of new mutations that are beneficial, deleterious or neutral (Eyre-Walker and Keightley 2007; Loewe and Hill 2010; Bataillon and Bailey 2014). It provides a snapshot of the robustness of the genome to changes in the DNA and carries information about the expected amount of genetic diversity within populations. Moreover, the beneficial part of the DFE informs on the adaptive potential of populations when introduced into a new environment (Sniegowski and Gerrish 2010; Bataillon and Bailey 2014). However, beneficial mutations in one environment can be deleterious in another, potentially resulting in so-called costs of adaptation (Bataillon et al. 2011), also termed antagonistic pleiotropy. So far, it has been difficult to address the prevalence of such costs of adaptation, because measuring the fitness of the same mutations across various environments is not straightforward. Specifically, previous studies using a selection of mutations obtained from laboratory evolution or mutation accumulation experiments found that antagonistic pleiotropy was rare (Ostrowski et al. 2005; Dillon et al. 2016; Sane et al. 2018). It is unknown whether this pattern holds for an unbiased selection of mutants.

Comparing the fitness of a large unbiased selection of mutants across environments has become feasible with the advancement of deep mutational scanning. Developed around a decade ago, deep mutational scanning allows for the assessment of the complete DFE of a focal genomic region in some genetically modifiable and fast-growing model species, using a combination of site-directed mutagenesis and deep sequencing (Fowler et al. 2010; Hietpas et al. 2011; Boucher et al. 2014; Logacheva et

al. 2016). Deep mutational scanning studies from single environments usually report a bimodal DFE with two peaks that represent neutral and strongly deleterious mutations, respectively (Hietpas et al. 2011; Acevedo et al. 2014; Boucher et al. 2014). Despite the ample use of this approach (Hietpas et al. 2011; Melamed et al. 2013; Acevedo et al. 2014; Boucher et al. 2014; Doud and Bloom 2016; Sarkisyan et al. 2016), few studies have quantified the impact of environmental challenges on the shape of these DFEs (but see (Hietpas et al. 2013; Bank et al. 2014). However, experimentally quantifying the DFE across environments is important because natural environments are constantly changing, which can have diverse evolutionary consequences that are directly related to the shape of the DFE (Dhar et al. 2011; Arribas et al. 2014; Mumby and van Woesik 2014; Brennan et al. 2017).

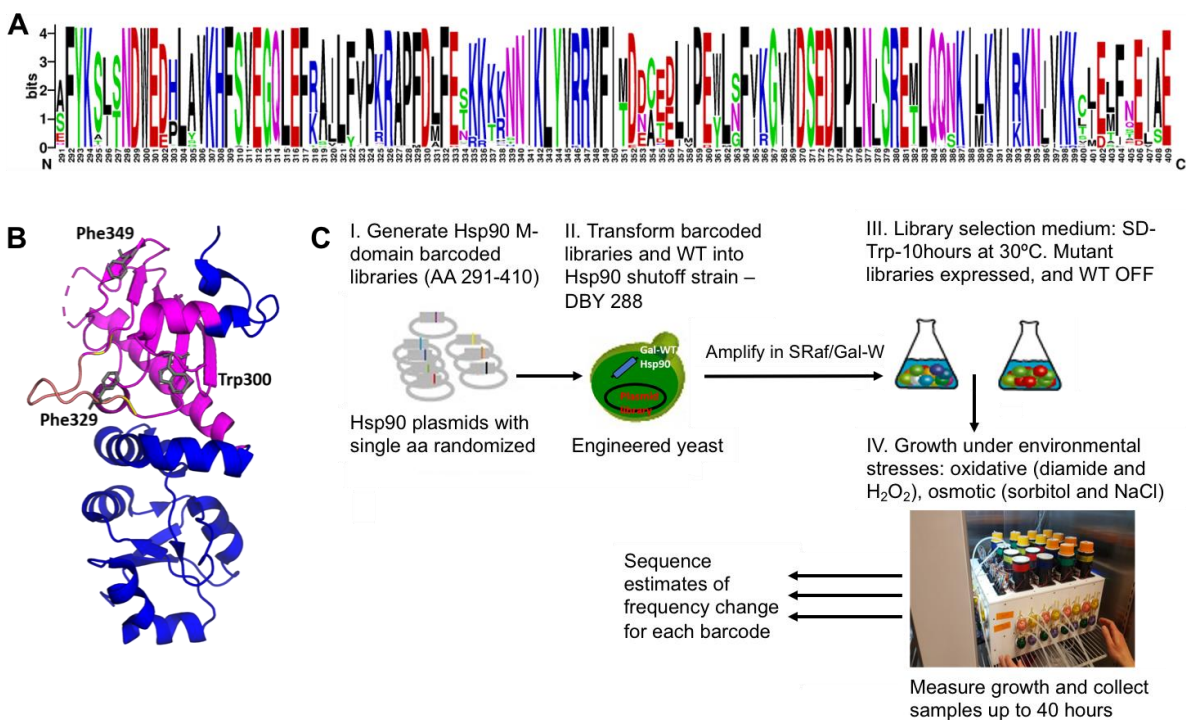
The consequences of environmental challenges on organisms manifest at many biological levels, including the protein level. For example, increased temperature can cause protein unfolding and aggregation, which can disrupt function and ultimately affect organismal fitness and survival (Richter et al. 2010). Chaperones, such as the heat shock protein Hsp90, help the cellular machinery survive stress conditions (Chen and Arnold 1993). By buffering deleterious fitness effects in stress conditions, chaperones are important for the short-term response to new and recurring environmental challenges. On a longer time scale, their buffering effect has also been argued to facilitate the maintenance of standing genetic variation elsewhere in the genome that can enable rapid genetic adaptation to new stress conditions (Rutherford 2003; Barrett and Schluter 2008; Jarosz and Lindquist 2010; Fitzgerald and Rosenberg 2019). Therefore, it is important to understand how chaperones evolve, and how the

selection pressure on a chaperone changes upon exposure to different environmental challenges.

The heat shock protein Hsp90 is a chaperone that plays an essential role in protecting cells from environmental stress and that is required at elevated levels for yeast growth at high temperature (Borkovich et al. 1989). Recent studies using systematic mutagenesis have begun to address how mutations to a strongly conserved client binding site of Hsp90 can impact evolutionary adaptation in yeast (Hietpas et al. 2013). Multiple mutations in a nine amino acid client-binding site of yeast Hsp90 provided a growth advantage under elevated salinity conditions (Hietpas et al. 2013; Bank et al. 2014). A recent larger-scale study found that at low Hsp90 expression levels, changes in environment greatly changed the shape of the DFE, and that some environments showed a higher prevalence of both beneficial and deleterious mutations (Flynn et al. 2020). However, another previous study proposed that at low expression the fitness effects of mutations, especially deleterious ones, should be larger (Jiang et al. 2013). Thus, it is unknown how much of the observed effect was due to expression level and how much was due to environmental changes.

Here we examined a 119 amino acid region (encompassing positions 291-409, Fig. 2.1) of the middle domain of yeast Hsp90 (aka Hsp82) at normal expression levels. Several studies have demonstrated that the middle domain of Hsp90 plays a prominent role in client binding (Nathan and Lindquist 1995; Nathan et al. 1997; Meyer et al. 2003; Hawle et al. 2006; Hagn et al. 2011), and suggested that mutations in this region may impact the relative affinity or priority of different clients and physiological pathways with

the potential to provide an adaptive benefit (Sato et al. 2000; Zhang et al. 2005; Hagn et al. 2011; Verba et al. 2016; Czemerer et al. 2017).



**Figure 2.1: Middle domain of Hsp90** (A) Amino acid conservation of yeast Hsp90 compared to 261 eukaryotic sequences. Relative height of the amino acid indicates degree of conservation. (B) Structural representation based on 1HK7.PDB of the middle domain of Hsp90, with amino acids 291-409 that are the focus of this study highlighted in purple. A solvent exposed amphipathic loop implicated in client binding (amino acids 327-341) is highlighted in yellow and residues implicated in client binding (amino acids W300, F329, F349) are shown as gray sticks. (C) Systematic approach to measure the adaptive potential of the middle domain of yeast Hsp90 (amino acids 291-409).

We used the EMPIRIC approach of deep mutational scanning (Hietpas et al. 2011) to estimate the selection coefficients of all amino-acid changing mutations in the middle domain of yeast Hsp90 across five environments. Using the inferred DFEs, we identified regions of the middle domain that stand out with respect to their potential for adaptation upon environmental change. Moreover, as some environments share the type of stress

that they induce, namely osmotic stress (0.5M salinity or 0.6M sorbitol) and oxidative stress (0.6 mM H<sub>2</sub>O<sub>2</sub> or 0.85mM diamide), we were able to study the impact of the type of stress on the adaptive potential of new mutations. To this end we: a) quantified the impact of environmental changes on the overall shape of the DFE, b) identified regions that showed the largest proportions of beneficial or deleterious mutations, respectively, c) quantified hotspots of costs of adaptation, and d) compared the identified regions with known client binding or other structurally important sites to connect the phenotypic and fitness effects of mutations. Altogether, we mapped potential protein regions that may play an important role in adaptation to different environments.

## **Results and Discussion**

To investigate the adaptive potential of the middle domain of yeast Hsp90 we used systematic site-directed mutagenesis of amino acid positions 291-409 that include known client binding sites. This resulted in ≈2300 amino-acid changing mutations, whose selection coefficients were estimated from 2-3 replicates of bulk competitions that were performed in five environments. We focused our analysis on standard lab conditions and four environmental stresses that affect growth rate (Supplementary Figure 2.1) in yeast (Gasch et al. 2000).

### **DFEs Across All Environments Show Many Wild-Type Like Mutations.**

The shape of the DFE indicates the relative importance of purifying or directional selection as compared with neutral evolution. In apparent contrast to the strong conservation of the Hsp90 middle domain in natural populations, we observed DFEs with mostly wild-type like mutations across all environments (Supplementary Figure 2.2,

2.3). Throughout the manuscript, we use the term “wild-type like” to denote mutations that are indistinguishable from the wild-type reference in the limit of experimental accuracy, see Material and Methods. We categorized mutants as wild-type like if the 95% confidence interval of the estimated selection coefficient overlapped with 0 (see also Supplementary Table 2.1, Materials & Methods). According to this criterion, between 50% (in the H<sub>2</sub>O<sub>2</sub> environment) and 65% (in the standard environment) of mutations showed a fitness effect that is indistinguishable from the reference type. Large numbers of wild-type like mutations have been observed previously in deep mutational scanning studies of DFEs (Soskine and Tawfik 2010; Hietpas et al. 2013; Melamed et al. 2013; Bank et al. 2014; Hom et al. 2019). Both biological and technical factors can be invoked as an explanation for the large number of wild-type like mutations. Firstly, the resolution of the experiment is likely much lower than the resolution at which natural selection may act in large yeast populations (Ohta 1992; Boucher et al. 2016). Secondly, selection pressures in the laboratory might differ greatly from those in nature (e.g., (Reznick and Ghalambor 2005; Kvitek and Sherlock 2013). Finally and relatedly, natural environments might be fundamentally different and rapidly fluctuating (Mustonen 2009). For example, a recent study of the DFE of the full Hsp90 sequence found that mutations which were tolerated across a set of diverse environments were those most likely observed in natural sequences (Flynn et al. 2020). We repeated the analysis of Flynn et al. 2020 and found the same pattern: our estimated DFEs across all environments of the subset of variants observed in nature (amino acid mutations observed across 267 eukaryotic sequences, see Material and Methods) show a high peak around 0. This results in a further enrichment of wild-type



like mutations in the subset of naturally observed variants as compared with the full data set (Supplementary Figure 2.5).

Our approach provides a dense scan of the local fitness landscape by measuring the selection coefficient of all amino-acid changing mutations that are available in a single mutational step from the ancestral state. The presence of a large number of wild-type like mutations in various different environments suggests that the local fitness landscape is rather flat, which is at odds with the strong conservation of the protein in yeast. However, further away from the wild type, epistatic interactions may condition the following mutational steps and thereby change the configuration of the fitness landscape, creating fewer attainable mutational paths and constraining evolution on longer time scales (e.g. (Weinreich et al. 2006; Kryazhimskiy et al. 2014).

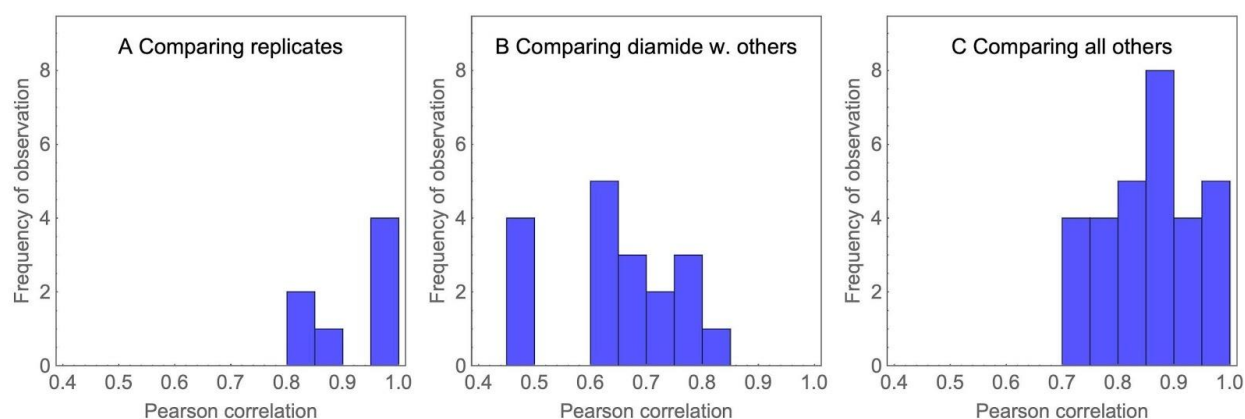
### **Correlations Across Conditions Reveal that Diamide Stands Out with Respect to Mutational Effects.**

We next quantified the correlations of fitness effects across replicates and environments (Figure 2.2, Supplementary Table 2.2). Consistent with the high accuracy of the experiment, we observed strong correlations of estimated mutational fitness effects between replicates (Supplementary Figure 2.6, mean Pearson correlation  $r=0.91$ ). Across pairs of environments, we observed a large variation of correlations ranging from  $r=0.48$  between diamide and salt to  $r=0.98$  between H<sub>2</sub>O<sub>2</sub> and the standard environment. In this analysis, diamide clearly stood out by showing consistently lower correlations of fitness effects with other environments than all others (mean correlation including diamide  $r=0.65$ , mean correlation of all others  $r=0.86$ ). This could be because diamide exerts multiple negative effects on the yeast cells. For

example, among the environments we investigated, diamide was the only condition that induces the expression of cell wall biosynthesis genes and genes involved in protein secretion and processing in the endoplasmic reticulum, which indicates its role in cell wall damage (Gasch et al. 2000). Diamide was also shown to affect individual transcription factors differently from H<sub>2</sub>O<sub>2</sub>, the other oxidative environment in our selection. One example that has been described in the literature is the Yap1 transcription factor, which is nuclear localized and active in diamide, but cytoplasmic and inactive in H<sub>2</sub>O<sub>2</sub> (Gulshan et al. 2011). Interactions of Hsp90 with other genes may also contribute to fitness differences of the mutations between the tested environmental conditions and mutants that we tested. Further global experimental analyses beyond the scope of this work will be required to determine the molecular features of diamide stress that elicit distinct selection on Hsp90 sequence.

Previous studies have reported that mutations had similar fitness effects across environments that shared metabolic features (Ostrowski et al. 2005; Dillon et al. 2016; Sane et al. 2018). Thus, we expected to see stronger correlations of fitness effects between environments that induce the same type of stress (salt vs sorbitol and H<sub>2</sub>O<sub>2</sub> vs diamide) than between different types of stress. Indeed, we observed a strong correlation of the fitness effects of mutations between salt and sorbitol ( $r= 0.79$ ) and between H<sub>2</sub>O<sub>2</sub> and diamide ( $r= 0.74$ ). However, these correlations are not much different from comparing any other pairs of environments (Supplementary Table 1). Again, diamide is the one environment that stands out with respect to this pattern. Here, the average correlation between diamide and other non-H<sub>2</sub>O<sub>2</sub> environments is 0.62 (vs  $r=0.74$  between H<sub>2</sub>O<sub>2</sub> and diamide), suggesting that, in this case, H<sub>2</sub>O<sub>2</sub> and diamide

may share some metabolic features where Hsp90 is involved. One possible explanation for this is that the oxidative stresses of diamide and  $\text{H}_2\text{O}_2$  cause an increase in the expression of chaperones, including Hsp90 (Gasch et al. 2000), inducing similar gene expression responses to those observed during thermal stress in yeast (Gasch et al. 2000).



**Figure 2.2: Histogram of correlations of fitness effects among replicates and environments reveals that diamide stands out as different.** The histograms correspond to comparisons of replicates (A), comparisons that include diamide (B), and all other comparisons (C). Correlations between replicates and between pairs of environments are in general high, whereas, correlations between diamide and the other environments tend to be lower. Altogether, this suggests that cells react in unique ways to the oxidative stress enacted by exposure to diamide, and that Hsp90 plays a specific role in the response to this stress.

We also observed a tendency for selection coefficients from the same batch of experiments to be more strongly correlated than selection coefficients inferred from different batches of experiments (Supplementary Table 2.2). This is not unusual (Venkataram et al. 2016) and may be due to small differences in the growth conditions between batches or during the sequencing steps. To mitigate this effect, we performed the analyses to detect beneficial and deleterious hotspots either based on at least two replicates or using the average of the estimated fitness effects between replicates (as

indicated below) and re-categorizing mutations based on the mutant category of both replicates (see also Materials and Methods).

The large correlations observed between most environments suggest that potential costs of adaptation (discussed in detail further below) should be the exception rather than the rule. This is at odds with previous results from a smaller region of Hsp90 (Hietpas et al. 2013), where most beneficial mutations detected at high salinity were deleterious in other environments. A potential explanation for this discrepancy lies in the choice of the protein region. Whereas, the previous study focused on only nine amino acids and thus chose a set of positions that are likely to show functionally specific patterns, we here had the experimental power to scan a larger region of the protein. Along a larger, less specifically chosen, stretch of the protein, it is likely that many positions and mutations play the same role across many environments. In this case, costs of adaptation of beneficial mutations could still be large, but due to their low proportion in the overall pool of observed mutations they barely affect the correlation of selection coefficients across environments. Indeed, when computing the correlation in a 10 amino acid region, we see high variability in the correlations (Supplementary Figure 2.7), e.g. with a range between 0.36 and 0.84 for the correlation between Diamide and Standard environments. This suggests that specific regions may represent functionally important positions (see below).

**Beneficial Mutations Are Present Across All Environments and Are Enriched in A Region of Hsp90 That Is Implicated in Stabilization of Interdomains and Client Binding Interfaces.**

Next, we identified the proportion and identity of putatively beneficial mutations across all treatments (Supplementary Table 2.1, Supplementary Figure 2.8, 2.9). We categorized mutants as putatively beneficial if the lower limit of the 95% confidence interval of the selection coefficient was larger than 0, and we considered as strongest candidates those that overlapped between replicates from the same environment. As reported above, wild-type like mutations are the most abundant category in all environments, with a low proportion but considerable number of beneficial mutations in almost all environments (Supplementary Figure 2.8, 2.9). We found the lowest number of beneficial mutations in H<sub>2</sub>O<sub>2</sub> and salt environments ( $n_{\text{H}_2\text{O}_2}=65$  and  $n_{\text{Salt}}=64$ , around 2.5% of the mutations) and the highest in diamide ( $n_{\text{Diamide}}=307$ , 12.3%).

The proportion of beneficial mutations, interpreted in the light of Fisher's Geometric Model (Fisher 1958; Tenaillon 2014), should be informative about the "harshness" of the experimental environments and the resulting potential for adaptation. Specifically, Fisher imagined that populations evolve in a multidimensional geometric phenotype space, where the same mutation is more likely to be beneficial if it happens in an individual that is far from the phenotypic optimum (Fisher 1958; Tenaillon 2014). Thus, we expected that in environments with large doubling times, corresponding to lower absolute fitness of the reference type (Supplementary Figure 2.1), we should observe a larger number of beneficial mutations. Moreover, we expected the lowest number of beneficials in the standard environment, to which the wild type is well adapted and in which it has the lowest doubling time. However, this expectation was met only in the diamide environment. At high salinity and in sorbitol, the observed proportion of beneficial mutations was lower than in the standard environment. This discordance with

expectations from FGM could be due to the model's very general assumptions which, for example, include that mutations affect all phenotypic dimensions with equal weight (see, e.g., (Harmand et al. 2017)). Many of these assumptions are likely violated here, because, for example, Hsp90 plays a very different functional role in the osmotic stress response in comparison to diamide induced oxidative stress - probably resulting in different phenotypic distributions of the same mutations in the (anyways rather abstract) phenotype space.

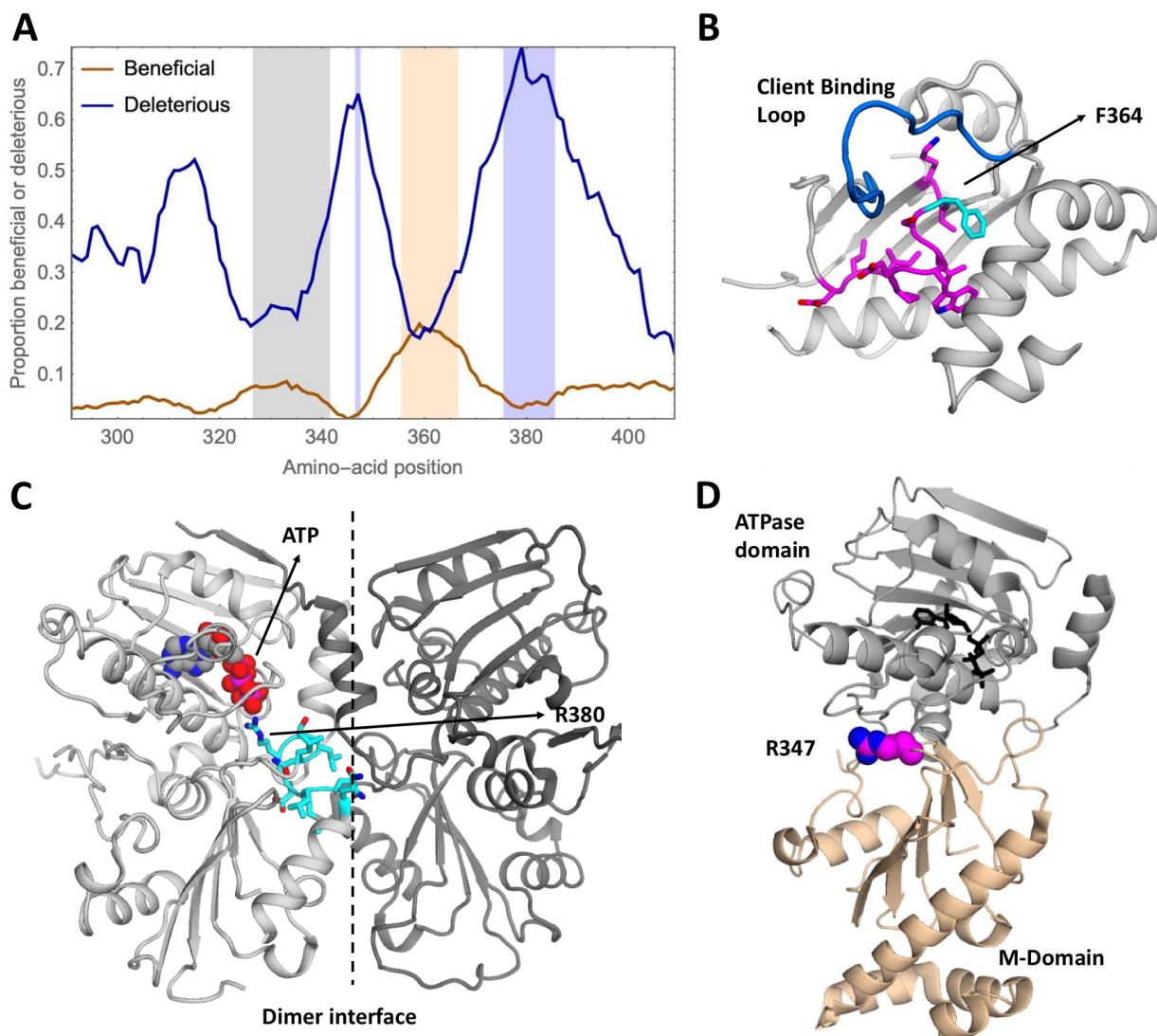
Specifically, it is known that Hsp90 basal function together with its co-chaperone Cdc37 is required for the induction of the osmotic stress response in yeast via activation of the Hog1 kinase in the HOG pathways (Hawle et al. 2007; Yang et al. 2007). However, increased salinity or sorbitol does not cause an increase in Hsp90 mRNA expression in comparison to diamide conditions, which cause an increase in mRNA expression similar to what is observed during heat shock (Gasch et al. 2000). The decoupling of function and expression, and the low number of beneficial mutations suggests that Hsp90s role during osmotic stress may be related to the activation of the general stress response mechanism of the cell (Mager and Siderius 2002). Another possibility is that osmotic stress does not cause heat shock induced protein unfolding (Richter et al. 2010) or diamide induced protein modifications that may compromise protein and cellular function (Gasch et al. 2000). Thus, under osmotic stress conditions the cell would not require increased expression of Hsp90. This suggests that mutations may play distinct roles for different conditions, perhaps under osmotic stress activation of clients and under diamide could be activation of clients and/or increase in Hsp90 expression and function.

Interestingly, beneficial mutations in diamide were dispersed across the whole middle domain (Supplementary Figure 2.8), whereas, most of the other environments showed a clear enrichment of beneficial mutations in the region of positions 356-366 (Figure 2.3, Supplementary Figure 2.8). This hotspot of beneficial mutations was particularly strong in sorbitol (Supplementary Figure 2.8). Out of the 31 beneficial mutations that showed a beneficial effect in at least two environments, 18 were located in this hotspot.

The beneficial hotspot contains residues that are part of an allosteric center which mediates distinct conformations in structures with and without client (Blacklock and Verkhivker 2013; Blacklock and Verkhivker 2014; Czemerer et al. 2017). Despite its different structural arrangements, the hydrophobic amino acids in this region are mostly buried from solvent (Figure 2.3B), indicating that the primary role of this region is to mediate the stability and rearrangement of different conformations. One position that particularly stands out with respect to the number of beneficial mutations is 364. Position 364 lies at the center of the cluster of the 10% largest average proportions of beneficial mutations (Figure 2.3A) and shows a clear deviation from the usually observed shape of the DFE (Supplementary Figure 2.10). Specifically, we identified 6 mutations at this position that are beneficial in at least two environments (Supplementary Figure 2.10). The burial of the large hydrophobic side chain of F364 should provide local stability. The beneficial mutations at position 364 and in this region in general suggest that disruption of local stability and conformational dynamics may alter Hsp90 function. Because this beneficial region partially overlaps with an allosteric center involved in Hsp90 conformational stability and dynamics (Blacklock and Verkhivker 2013; Blacklock and Verkhivker 2014) we hypothesize that Hsp90 function

may be altered by the identified candidate beneficial mutations in a manner that disrupts local stability and dynamics of this region. Furthermore, these disruptions may promote conformational changes in the neighboring client binding loop that allow this region to sample a larger and/or different conformational space in a manner that changes the relative affinity and thereby priority of different clients - a property that we speculate could provide benefits in specific conditions. Consistent with this hypothesis, the beneficial hotspot is adjacent to a known client binding loop (Figure 2.3) such that alterations in the structure or dynamics of the hotspot are likely to influence client binding nearby. It is possible that natural selection balances client priorities integrated over multiple conditions, which provides opportunities for Hsp90 mutations to improve priorities for individual conditions (Flynn et al. 2020).





**Figure 2.3: Hotspots of beneficial and deleterious mutations in the middle domain do not coincide with known client binding loop.** A) Average proportion of beneficial and deleterious mutations along the genome across all environments, illustrated using a 10-amino-acid sliding window. The region with the 10% largest proportions of beneficial mutations is highlighted in light orange (position 356-366), the region with the 10% largest proportions of deleterious mutations is highlighted in light blue (positions 347 and 376-385). A known client binding loop (positions 327-341) is highlighted in light gray. To avoid biases, the analysis was restricted to subsets of two replicates of all normal-expression environments. B) Structural representation of the Hsp90 middle domain (1HK7.PDB) illustrating the beneficial hotspot in magenta with residue 364 highlighted in cyan. Biochemical, structural and mutational studies identified positions 327-341 as a client binding loop, which is shown in blue. The beneficial hotspot that was identified in our analyses is adjacent to the above-mentioned client binding loop. C) Structural representation of the ATPase and middle domains of Hsp90 from 2CG9.PDB.

The deleterious hotspot (in cyan) contains a catalytic amino acid and is located at a dimerization interface. It includes residue R380 that stabilizes the leaving phosphate during ATP hydrolysis. Subunits are distinguished with different shading. D) The deleterious hotspot at position 347 is also located at the interface of the ATPase domain and the middle domain and is highlighted in pink and blue.

## **The Proportion of Deleterious, Wild-Type Like, and Beneficial Mutations Varies**

### **Greatly Along the Protein Sequence.**

We showed above that the overall correlations of fitness effects across environments were generally large, which indicates similar effects of the same mutations across environments. In contrast and similar to the local correlations of fitness effects (Supplementary Figure 2.7), the proportions of beneficial, wild-type like, and deleterious mutations vary greatly along the protein sequence. Whereas, the overall pattern is similar between environments, the relative proportions differ between environments (Supplementary Figure 2.8, 2.9). Our finding that the same positions are enriched for deleterious and beneficial mutations across environments suggests that the structural properties of the middle domain in these regions, rather than its binding partners, might be the most important factor for predicting the fitness effects of mutations.

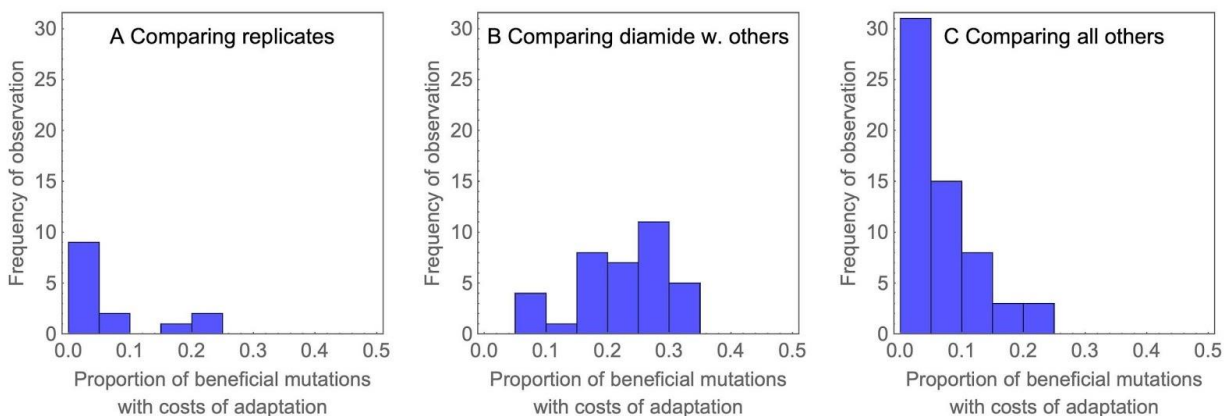
Interestingly, the regions with the largest proportions of deleterious mutations are located near the ends of the beneficial hotspot region (position 347 and 376-385, respectively) (Supplementary Figure 2.8). At the latter deleterious hotspot, around 70% of all mutations are deleterious in a sliding window of 10 amino-acid positions. Such a shift of the DFE towards deleterious mutations suggests that this protein region is under strong purifying selection. Structurally, the region 376-385 is part of a catalytic loop

required for ATP hydrolysis which is necessary for the activation of all known clients (Wolmarans et al. 2016). Position R380 in the catalytic loop binds to and stabilizes the leaving phosphate of ATP (Figure 2.3C) and mutations at this position compromise Hsp90 function and cell viability (Meyer et al. 2003). The efficiency of catalysts depends strongly on geometry, and the precise location of R380 relative to ATP is likely tightly linked to Hsp90 function. The regions adjacent to R380 appear to be important for positioning the catalytic arginine, providing a rationale for the strong purifying selection that we infer. The second deleterious hotspot at position 347 is also located at the interface of the ATPase domain and the middle domain (Figure 2.3D), consistent with current understanding that Hsp90 mechanism requires precise ATP-dependent interactions between these domains (Schopf et al. 2017).

### **Few Mutations Show Costs of Adaptation.**

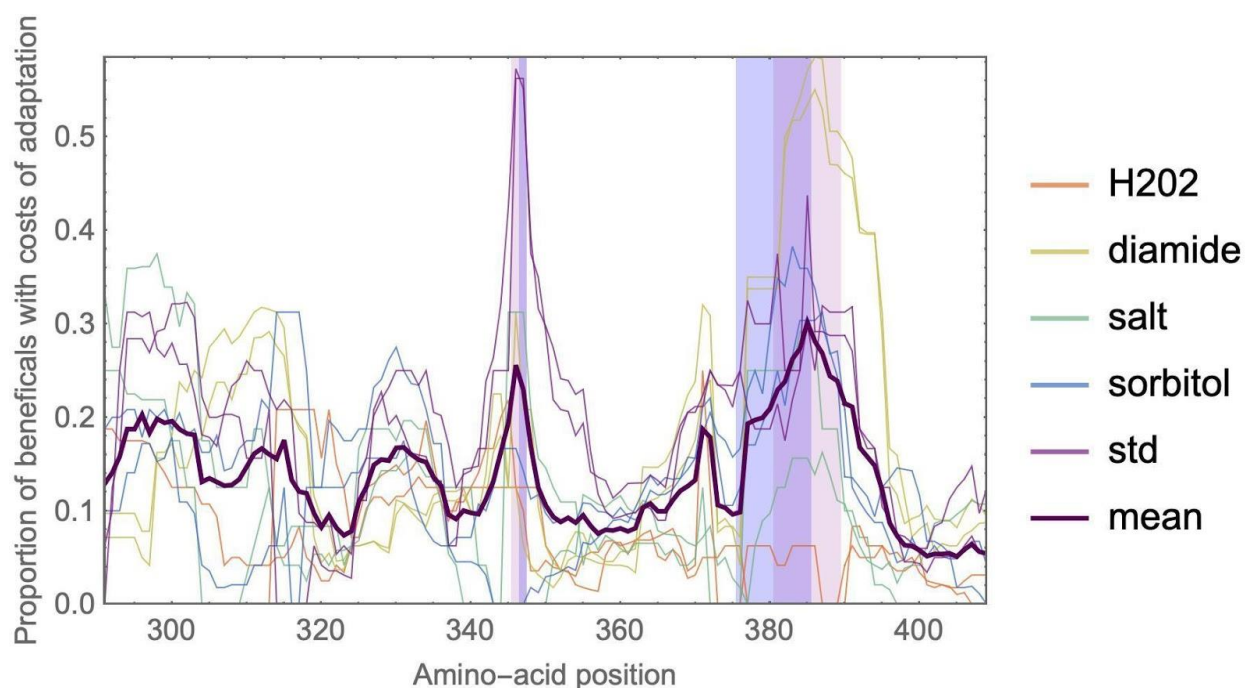
Among the identified beneficial mutations, we were specifically interested in those that are deleterious in other environments, which results in a so-called "cost of adaptation". The same phenomenon has also been termed antagonistic pleiotropy. In previous work we reported a large prevalence of costs of adaptation in a 9-amino-acid region of Hsp90 (Hietpas et al. 2013). In this study, we found that costs of adaptation were not pervasive (Figure 2.4, Supplementary Table 2.3). Consistent with the comparatively low correlations of fitness effects and the special role of the diamide stress discussed above, we observed the largest proportion of mutations that show costs of adaptation between diamide and other environments. Whereas, the mean proportion of mutations displaying costs of adaptation across all comparisons of

environments was 14.3%, comparisons that included diamide showed on average 22.7% mutations with costs of adaptation.



**Figure 2.4: Histogram of proportions of mutations that display costs of adaptation illustrates that costs of adaptation are generally rare, and most likely if comparisons involve diamide environments.** The histograms correspond to comparisons of replicates (A), the subset of other comparisons that include diamide (B), and all other comparisons (C).

Mapping the proportion of beneficial mutations that are deleterious in other environments along the protein sequence, we observed that there is a hotspot for costs of adaptation at amino-acid positions 381-391 (Figure 2. 5). Interestingly, this region showed costs of adaptation across various environments, which indicates that each environment has specific beneficial mutations which are deleterious in other environments. Structurally, this region also belongs to the catalytic loop involved in ATP hydrolysis discussed above. Indeed, the identified region partly overlaps with the above-discussed hotspot of deleterious mutations (positions 376-385), which is unsurprising since a larger proportion of deleterious mutations also increases the (technical) proportion of mutations that can be classified as costly.



**Figure 2.5: A hotspot of costs of adaptation is located at positions 381-391.** The 10% positions with the largest mean proportion of mutations that display costs of adaptation (dark purple line) are highlighted in light purple. This region greatly overlaps with the hotspots of deleterious mutations from Figure 3 (highlighted in blue). Thin curves indicate the mean proportion of beneficial mutations in a focal environment that is deleterious in another environment. A 10-amino-acid sliding window was used to locate and display region-specific effects. To avoid biases, the analysis was restricted to subsets of two replicates of all environments.

We next computed whether there was a correlation between the effects of the subset of beneficial mutations in one environment with their effect in the other environments. Indeed, the effects of beneficial mutations in diamide were negatively correlated with the same mutation's effect in all other environments (Figure 2.5, Supplementary Figure 2.11) except H<sub>2</sub>O<sub>2</sub>. In other words, the more beneficial a mutation was in diamide, the more deleterious it tended to be in salt, standard and sorbitol. Again, this suggests that the beneficial mutations we identified in diamide may be involved in the response to a very specific type of stress. Across all other environments,

we did not see any evidence that stronger beneficial mutations tended to have a more deleterious effect in other environments. In fact, there is a suggestive positive correlation between fitness effects of beneficial mutations with their respective effect in all other non-diamide environments, which points to the presence of synergistic pleiotropy. While generally defined as mutations having the same effect on more than one trait, in the context of this study synergistic pleiotropy would mean that beneficial mutations tend to have a similar ranking across environments. This is in line with results from studies of *E. coli* using experimental evolution (Ostrowski et al. 2005; Dillon et al. 2016) or mutation accumulation (Sane et al. 2018). These studies reported synergistic pleiotropy between different carbon source environments, and rare presence of antagonistic pleiotropy. Specifically, Sane et al (Sane et al. 2018) found that this pattern was maintained also for the categories of neutral and deleterious mutations, and that mutations with larger effect were more likely to show antagonistic pleiotropy. The reported prevalence of synergistic pleiotropy was associated with a sharing of the metabolism and transport of resources (Ostrowski et al. 2005; Dillon et al. 2016; Sane et al. 2018).

## **Conclusion**

Recent advances in experimental and technological approaches have led to the feasibility of large-scale screens of the DFE of new mutations, which, from an evolutionary point of view, is a key entity to determine the potential for adaptation. Here, we took a new route by mapping the proportions of beneficial, wild-type like, and deleterious mutations along a 119-amino-acid region of the Hsp90 protein in yeast, a protein that is heavily involved in the response to environmental stressors. This

approach allowed us to create a genotype-phenotype map to identify specific protein regions that may be important for adaptation to new environments. Specifically, by comparing the DFE along the protein sequence and between environments, we identified hotspots of beneficial and deleterious mutations that are shared between environments, and a region in which beneficial mutations in one environment tend to be deleterious in other environments. Interestingly, neither of these regions coincided with the best described client binding loop in the studied region, which we had *a priori* considered the most likely candidate to display patterns different from the rest of the region. Moreover, our analyses suggested that mutational effects generally differ little across environments except in diamide, which stood out both with respect to the number and also the distribution of beneficial mutations along the studied protein region. Altogether, our study of the DFE across environments sheds light on the evolutionary role of a specific protein region from a new perspective.

## **Materials and Methods**

### **Generating Point Mutations.**

To accurately measure the fitness effects of all possible point mutations in a large portion of the middle (client-binding) domain of yeast Hsp90, we used saturation mutagenesis at positions 291-409. We used a cassette ligation strategy to introduce mutations as previously described (Hietpas et al. 2012). This strategy reduces the likelihood of secondary mutations because it avoids the potential for errors during PCR steps. As a control for the mutational procedure, twelve positions were randomized in isolation and Sanger sequenced to assess the level of incorporation of all four nucleotides at each position in the target codon. At all randomized nucleotide positions

within these twelve samples, we observed a similar magnitude of signal for each of the four nucleotides. For larger scale production, libraries were generated at 10 positions at a time as previously described (Hietpas et al. 2012). As additional controls, we generated a sample containing individual stop codons as well as the parental wild-type Hsp90 sequence. All variants of Hsp90 were generated in a plasmid (pRS414GPD) previously shown to produce endogenous levels of Hsp90 protein in yeast (Chang and Lindquist 1994).

### **Constructing Barcoded Libraries.**

To improve the efficiency and accuracy of fitness estimates, we added barcodes to a non-functional region of the plasmid, ~200 NTs downstream from the 3' untranslated region of Hsp90. Barcodes were introduced using a cassette ligation strategy (Supplementary Figure 2.12). Plasmid libraries were treated with NotI and Ascl to generate directional sticky ends. NotI and Ascl recognize 8-base cut sites that are unlikely to cut any of the Hsp90 variants in the library. We designed and annealed barcode forward and barcode reverse oligos together such that the resulting duplex product included a central N18 region bracketed by constant regions that facilitate annealing and overhangs that direct directional ligation into NotI and Ascl overhangs. One of the constant ends in the designed oligo cassette contains an annealing region for an Illumina sequencing primer. Barcoded libraries were transformed into *E. coli* and pooled into a bulk culture that contained about 10-fold more transformants than Hsp90 variants in the library. We purified barcoded plasmids from this bulk culture. This procedure resulted in approximately 20 barcodes for each Hsp90 codon variant in the library (Supplementary Figure 2.13). The potential diversity in the



N18 barcode that we used ( $4^{18} \sim 10^{11}$ ) far exceeds the number of barcodes that we utilize ( $\sim 64 \times 119 \times 10^5$ ), which makes it likely that each Hsp90 variant will have a barcode that differs from all other barcodes at multiple nucleotides. With this setup, errors in sequencing of barcodes can be detected and eliminated from further analysis, which reduces the impact of sequencing misreads on estimates of variant frequency and fitness. Additional controls consisting of individual stop codons and wildtype Hsp90 were barcoded separately. For these controls, we isolated barcoded plasmid DNA from individual bacterial colonies and determined the barcodes by Sanger sequencing (Supplementary Figure 2.12).

### **Associating Barcodes to Mutants.**

To identify the barcodes associated with each Hsp90 variant in our libraries, we used a paired-end sequencing approach essentially as previously described (Hiatt et al. 2010), see also Supplementary Figure 2.14). Using paired-end sequencing on an Illumina MiSeq Instrument barcodes were associated with variant genotypes via reading from the Hsp90 gene with a 250 base-pair read and the associated N18 barcode with a 50 base-pair read. To reduce the size of the DNA fragments for efficient Illumina sequencing, we removed a portion of the Hsp90 gene such that the randomized regions were closer to the N18 barcode. This was done to increase the density of DNA on the sequencer by reducing the radius of clonal clusters during sequencing. Plasmid DNA was linearized with *StuI* and *NotI* endonucleases that removed  $\sim 400$  bp's of DNA. The linearized products were circularized by blunt ending followed by ligation at DNA concentrations that predominantly lead to unimolecular ligations (Revie et al. 1988). The resulting DNA products were amplified using a PE2\_F primer and the standard Illumina

PE1 primer that anneals next to the N18 barcode. Two PE2\_F primers were designed in order to read across the region of Hsp90 that we randomized. PCR products were gel purified and submitted for paired end sequencing. We obtained sufficient paired-end reads such that the average barcode was read more than 10 times (Supplementary figure 2.13). The paired-end sequencing data was subjected to a custom data analysis pipeline to associate Hsp90 variants with barcodes. First, very low-quality reads with any Phred score less than 10 in reads 1 or 2 were discarded. Next, the data were organized by the barcode sequence. For barcodes with three or more reads, we constructed a consensus of the Hsp90 sequence read. We compared the consensus sequence to the parental Hsp90 sequence in order to determine mutations. Consensus sequences containing more than one protein mutation were discarded. The pipeline output generated a file organized by point mutation that lists all barcodes associated with that mutation (Supplementary Data). This file was used as the basis for calling a variant based on barcode reads.

### **Yeast Competitions.**

As in previous work (Hietpas et al. 2012; Hietpas et al. 2013), we performed Hsp90 competitions using a shutoff strain of *S. cerevisiae* (DBY288) (Supplementary Figure 2.14). The sole copy of Hsp90 in DBY288 is driven by a galactose-inducible promoter, such that the strain requires galactose for viability and cannot grow on glucose. The introduction of a functional Hsp90 variant driven by a constitutive promoter rescues the growth of DBY288 in glucose media. We introduced the library of middle domain variants of Hsp90 driven by a constitutive promoter into DBY288 cells using the lithium acetate method (Gietz and Schiestl 2007). The efficiency of transformation was more

than 10-fold higher than the number of barcodes, such that the average barcode was transformed into more than 10 individual cells. To enable the analyses of full biological replicates, transformations were performed in replicate such that a separate population of yeast were transformed for each biological replicate. Transformed yeast were initially selected in RGal-W (synthetic media that lacked tryptophan and contained 1% raffinose and 1% galactose supplemented with 100 µg/mL ampicillin to hinder bacterial contamination) to select for the plasmid, but not function of the plasmid encoded Hsp90 variant. This enabled us to generate a yeast library containing Hsp90 variants that could support a full range of fitness from null to adaptive. Cells were grown for 48 hours at 30 °C in liquid SRGal-W until the culture was visibly opaque compared to a control transformation sample that lacked plasmid DNA but was otherwise identical to the library sample.

To initiate the competition, cells were transferred to shut-off conditions and different environmental stresses. In order to deplete the pool of wild type Hsp90 protein, the library sample was diluted into SD-W (synthetic media lacking tryptophan with 2% glucose and 100 µg/mL of ampicillin) and grown at 30 °C for 10 hours. After depletion of wild type Hsp90 protein, the cells were split into different stress conditions in SD-W media at 30 °C: salt stress (0.5M NaCl), osmotic stress (0.6M sorbitol), oxidative stress (0.6 mM hydrogen peroxide or 0.85 mM diamide), as well as a control non-stress condition. We used a custom built turbidostat (Supplementary Figure 2.14) to provide constant growth conditions during this competition phase of the experiment. Cells were grown under consistent density and population size.  $10^9$  cells were maintained in a 50 mL volume over the course of 40 hours for each condition. Rapid magnetic stirring was

used to provide aeration to the media and cells. Samples of  $10^8$  cells were collected after 0, 4, 8, 12, 24, 32, 40, and 48 hours of competition in the different conditions. At the time of collection, samples were centrifuged and the pellets immediately frozen at  $-80\text{ }^{\circ}\text{C}$ .

### **Sequencing of Competition Samples.**

The fitness effects of Hsp90 variants were estimated based on frequencies observed in next-generation sequencing analyses essentially as previously described (Hietpas et al. 2012). DNA was isolated from each timepoint sample as described (Hietpas et al. 2013) and the barcodes were amplified and sequenced. Barcodes were amplified using the standard Illumina PE1 primer that anneals next to the N18 barcode and custom designed barcode\_forward primers. A set of barcode\_forward primers were designed with identifier sequences that could be read during sequencing and used to distinguish each timepoint sample (Supplementary Figure 2.14C). Each identifier sequence was eight nucleotides in length and differed by at least two bases from all other identifier sequences. Twenty cycles of PCR were sufficient to generate clear products from all samples. These PCR products were purified on silica columns (Zymo Research) and sequenced for 100 base single reads on an Illumina NextSeq instrument. The barcode corresponded to the first 18 bases and the identifier at positions 91-98. The resulting fastq files were processed and analyzed using customized software tools. First, poor quality reads containing any positions with a Phred score less than 20 were discarded. Reads were tabulated if the barcode matched a barcode associated with a point mutation and if the identifier matched with a timepoint. The analyses scripts output a file with the number of reads for each amino acid point mutant at each timepoint in each

condition (Supplementary Data). Barcodes with 0 reads at the first time point, or with a total of 1 read along the whole trajectory were removed from the analysis.

### **Estimation of Selection Coefficients.**

Inference of selection coefficients was performed via log-linear regression as described in (Matuszewski et al. 2016) (see also Supplementary Material). To improve estimation accuracy by incorporating information from each individual barcode, the linear model for each amino-acid changing mutation included barcode identities as nominal variables. For each amino-acid changing mutation we obtained a selection coefficient and 95% confidence interval (CI) of the estimate, representing the variation within amino acid due to differences in codon\*Barcode tag throughout time. This reduces the impact of potential outliers and averages over synonymous codons within amino acid. Mutants with 50 or less total reads at the first time point were removed from the data set. Finally, we normalized all selection coefficients by subtracting the median of all mutations that were synonymous to the wild type. This ensures that the average of the selection coefficients of wild type synonyms represents a selection coefficient of 0. To categorize mutations as beneficial, deleterious or wild-type we tested the overlap of the CI with 0. Namely, if the lower CI limit was larger than 0, a mutation was considered beneficial, and if the upper CI limit was smaller than 0, a mutation was considered deleterious. All other mutations were considered wild-type like. We use this terminology to distinguish between neutral and wild-type like mutations. Specifically, we expect that at large population sizes the number of wild-type like mutations is larger than the number of neutral mutations. This is because neutral mutations should behave neutrally with respect to the neutral or nearly-neutral theories, i.e., when  $2N_s < 1$  (Ohta 1992). This

threshold for neutrality is far below the measurement accuracy of our experimental setup. Thus, we instead consider as "wild-type like" those mutations that are indistinguishable from the wild type according to our inference accuracy.

### **Natural Variants:**

We identified amino acid mutations from Hsp90 sequences of 261 eukaryotic organisms (Starr et al. 2018). Then, for each environment, we compared the distribution of fitness effects of the identified mutations (natural variants) with the DFE from all mutations. Finally, we also computed the overall proportion of beneficial, deleterious and wildtype like mutations across environments and for the natural variants.

### **Costs of Adaptation.**

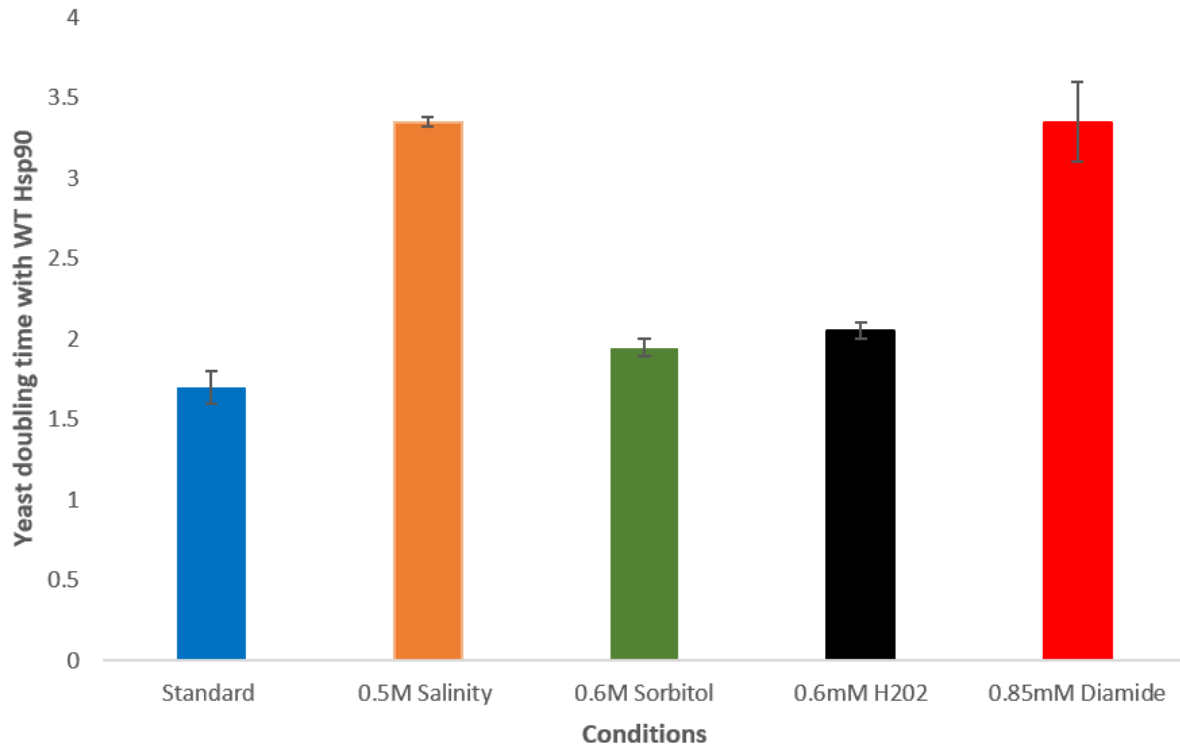
To compute the costs of adaptation for Figure 2.4 and 2.5, we first selected all mutations in a focal environment that were categorized as beneficial. We then computed the proportion of this subset of mutations that was deleterious either in one other environment (for Figure 2.4) or across all other environments (excluding replicates of the focal environment for Figure 2.5). Calculating the proportion of mutations with costs of adaptation compensates for variable numbers of beneficial mutations across environments, which may occur due to both biological and experimental reasons (e.g., different error margins). For Figure 2.2 and 2.4, all available data sets were considered for the analysis. For Figure 2.3 and 2.5, only two replicates of each environment were considered. That was done in order to avoid biases due to the under- or overrepresentation of environments in the analysis.

Analyses were performed using R version 3.5.1 (Team. 2018), Mathematica 12.0.0.0, Python 2.7.12 and Pymol 0.99. Code is available as Supplementary Material.

All data will be made available in a data repository upon acceptance of the manuscript.

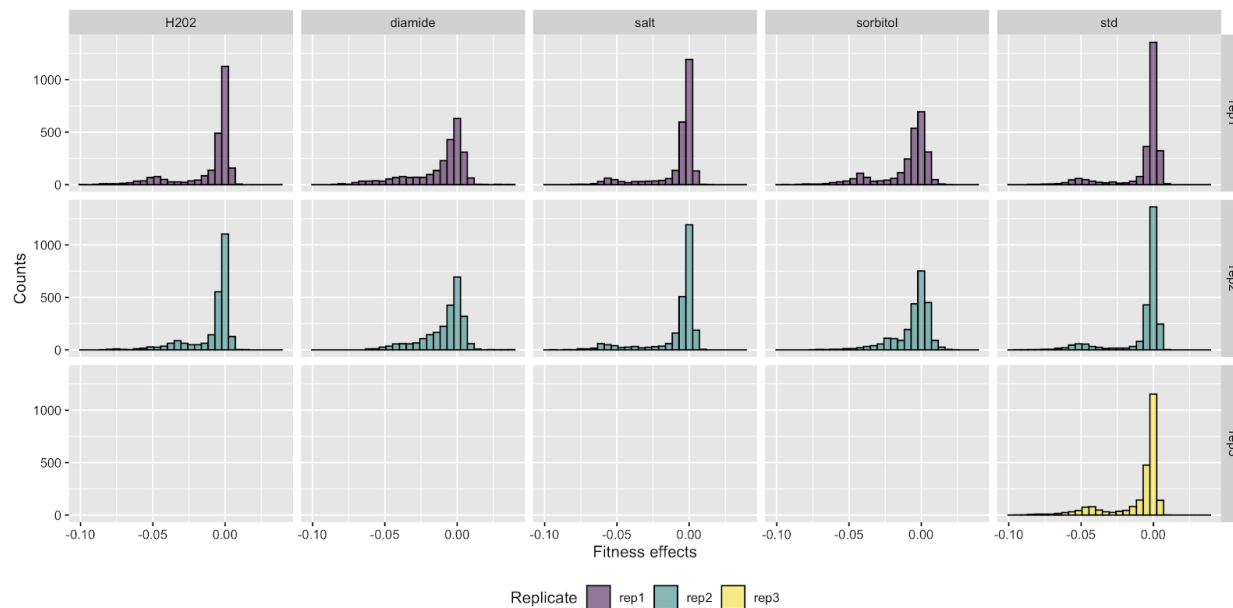
### **Acknowledgements**

This work was supported in part by grant R01GM112844 from the National Institutes of Health (to DNAB). IF was supported by a postdoctoral fellowship from the FCT (Fundação para a Ciência e a Tecnologia) within the project JPIAMR/0001/2016. CB is grateful for support by EMBO Installation Grant IG4152 and by ERC Starting Grant 804569 - FIT2GO.



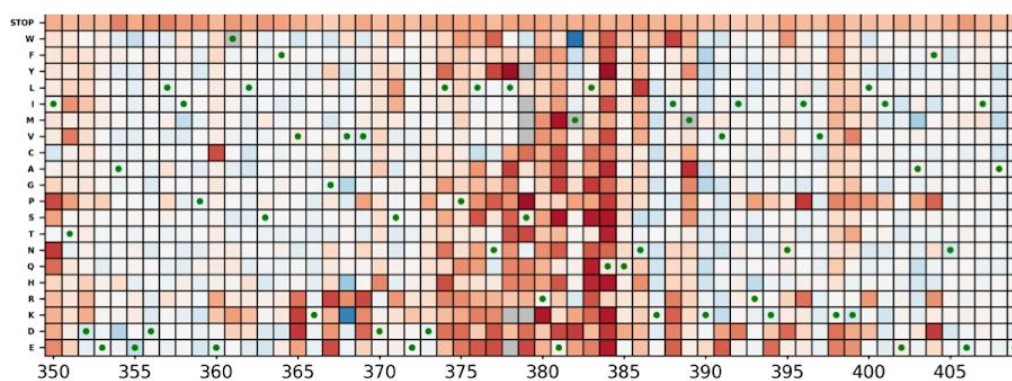
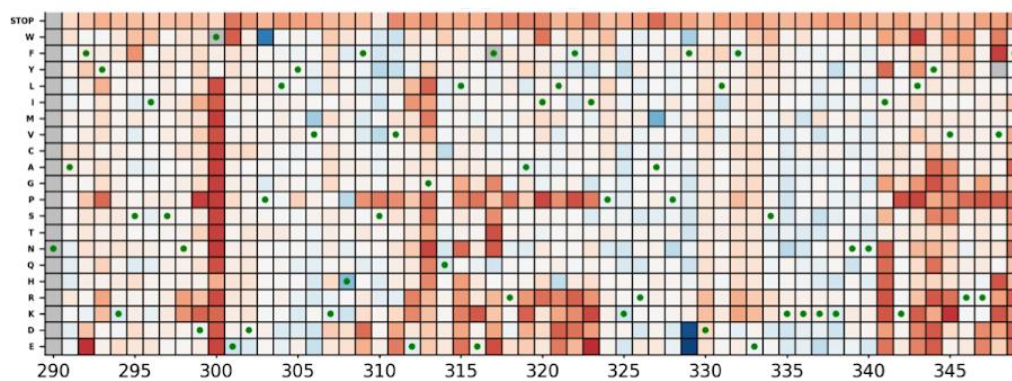
**Supplementary Figure 2.1. Doubling times for WT Hsp90 yeast growth under all environments.** In general, there is an increase in the doubling time in all environments tested compared to the standard environment. This increase is larger in high salinity and diamide environments, suggesting that these environments should present the highest stress load to the protein.



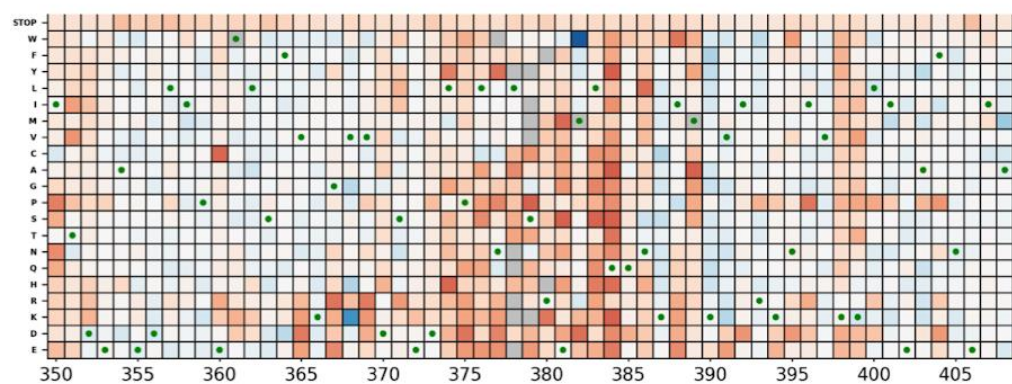
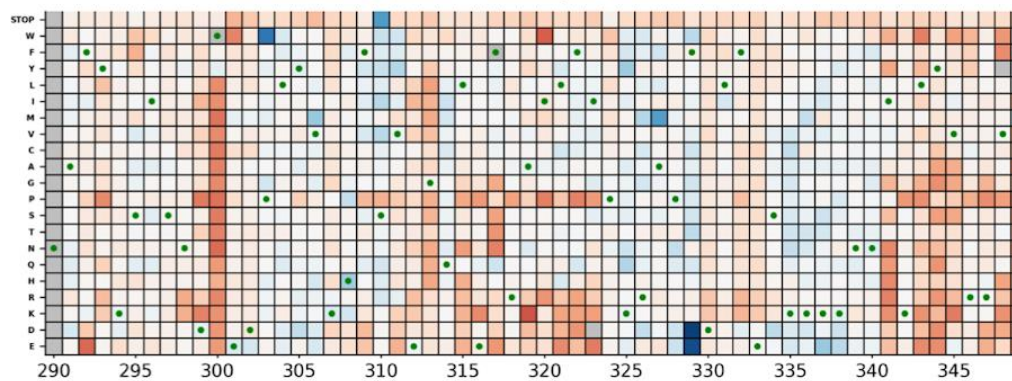


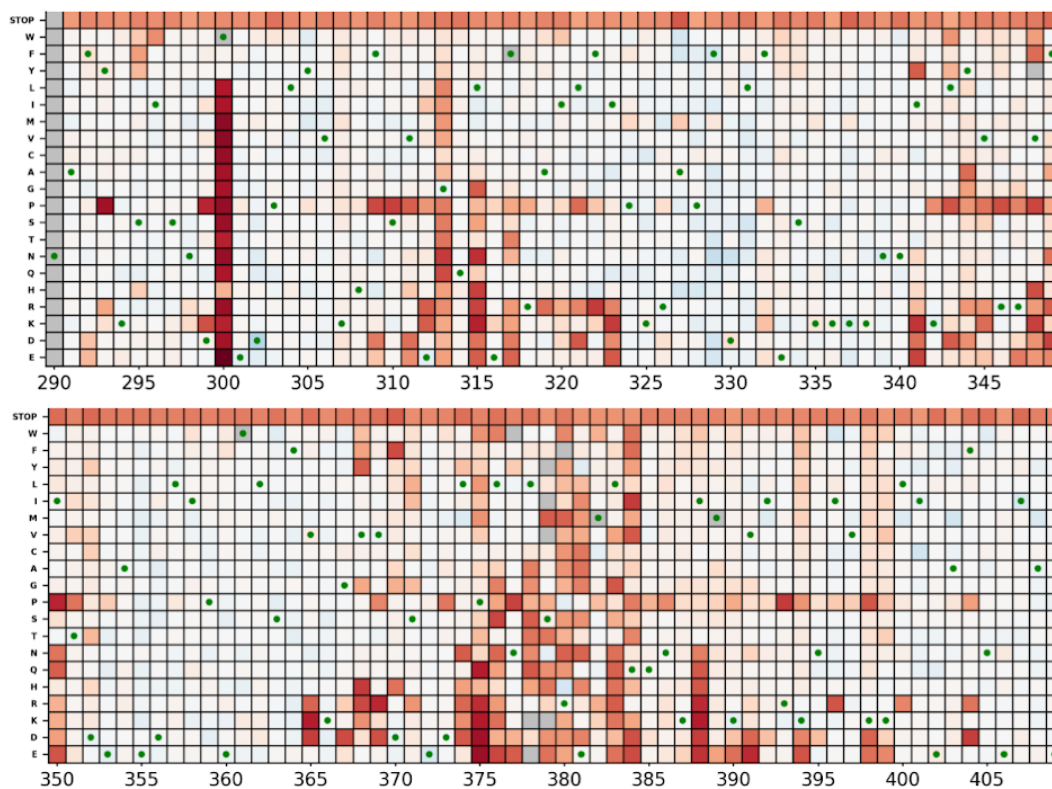
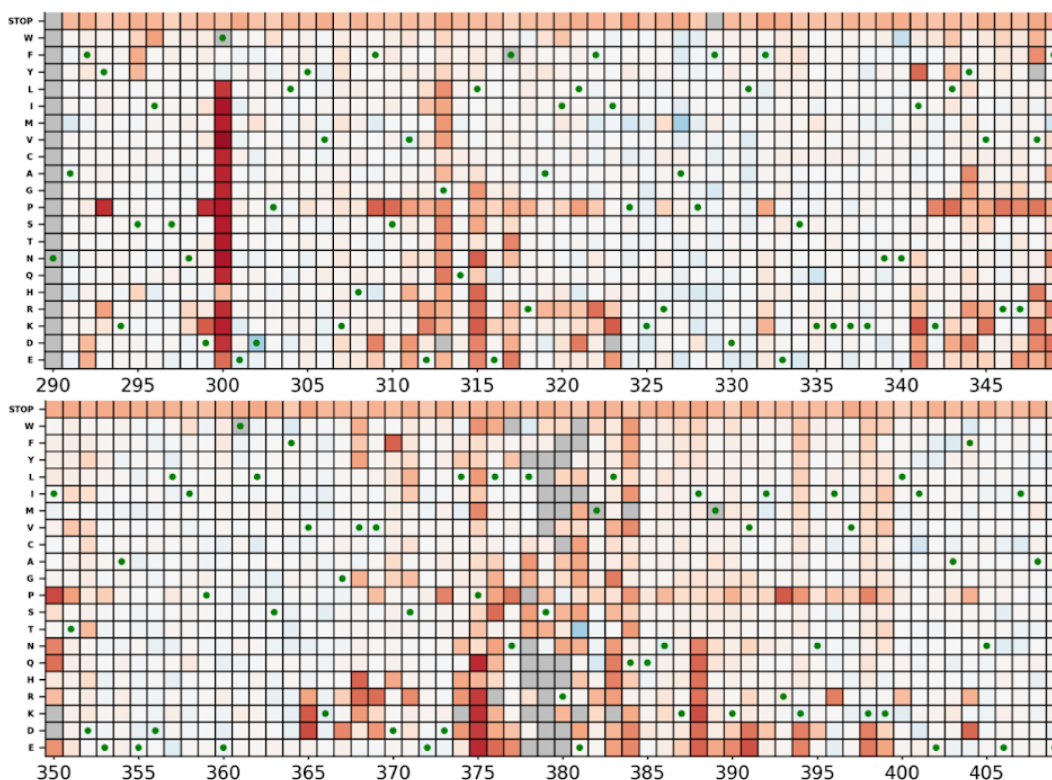
**Supplementary Figure 2.2: Distribution of Fitness Effects (DFE) across environments.** Replicates within an environment largely overlap. There is a marked difference of the DFE in diamide compared to the other environments. Interestingly, DFEs in H<sub>2</sub>O<sub>2</sub> and salt show a similar shape to the DFE in the standard environment, with a large number of mutations in the wild-type like region and a smaller peak of deleterious mutations, at which stop codons are located. In diamide and sorbitol, the DFE shows a lower number of wild-type like mutants, a larger variance, and a higher proportion of weakly and strongly deleterious mutations.

## Diamide Replicate 1

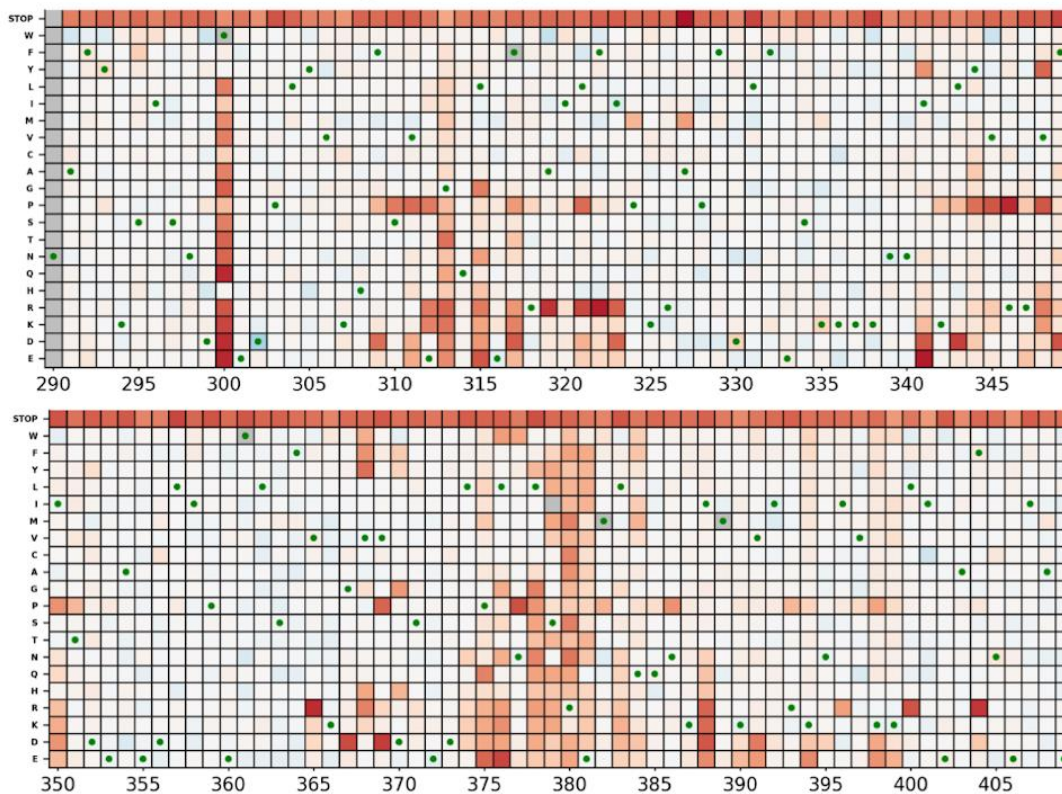


## Diamide Replicate 2

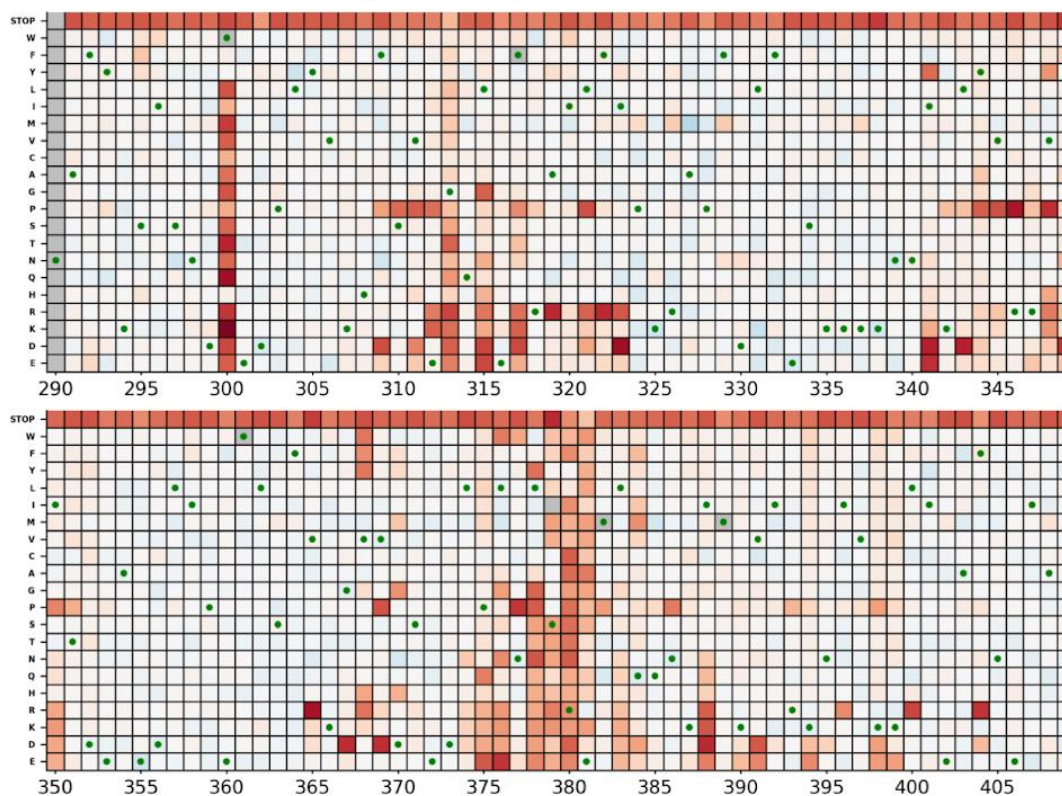


H<sub>2</sub>O<sub>2</sub> Replicate 1H<sub>2</sub>O<sub>2</sub> Replicate 2

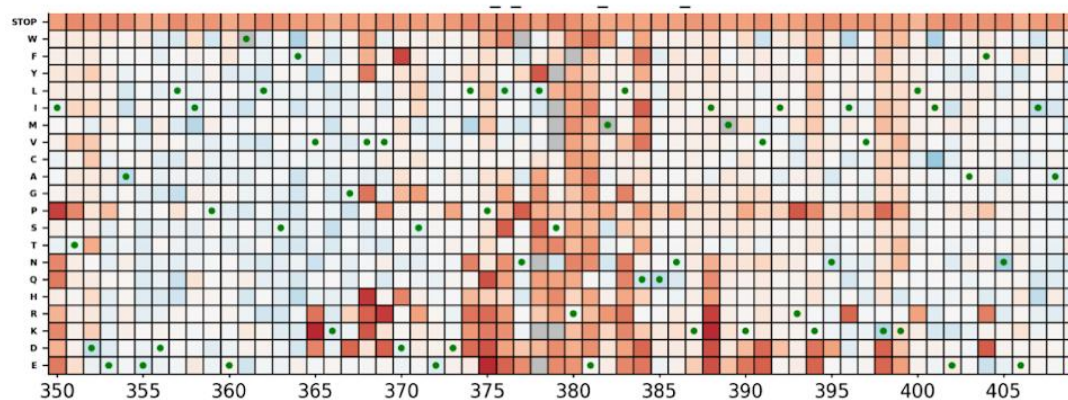
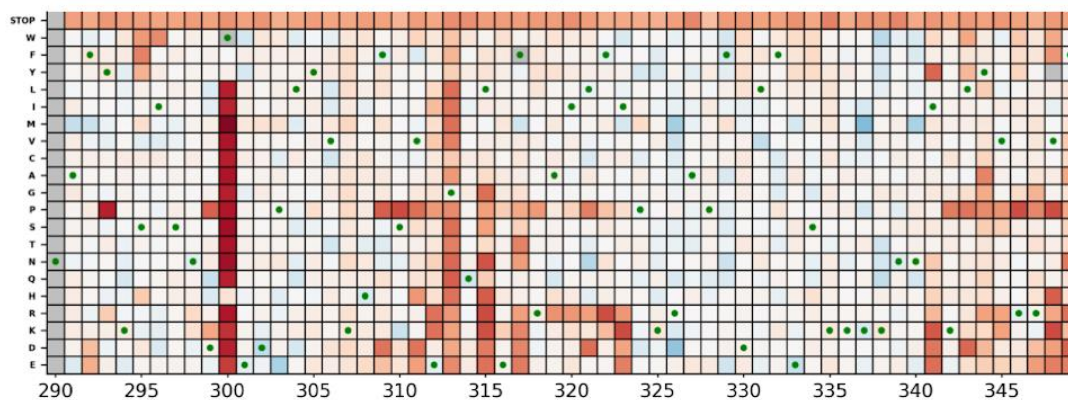
## High Salinity Replicate 1



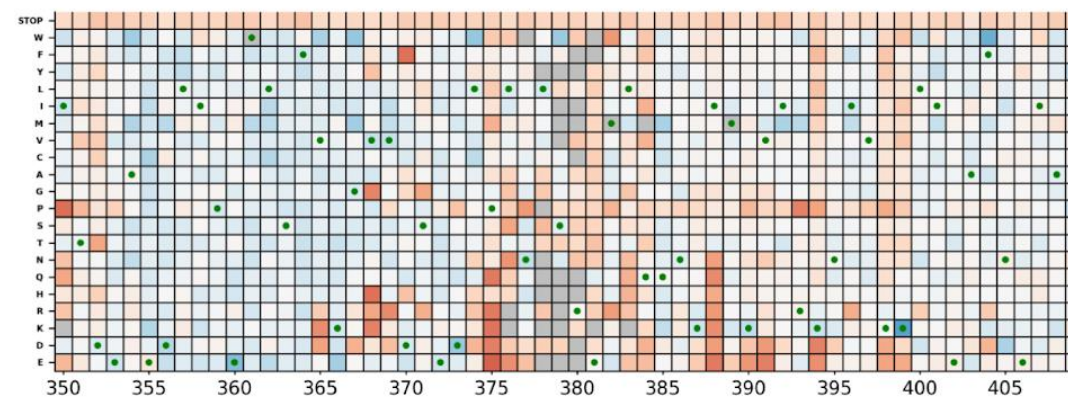
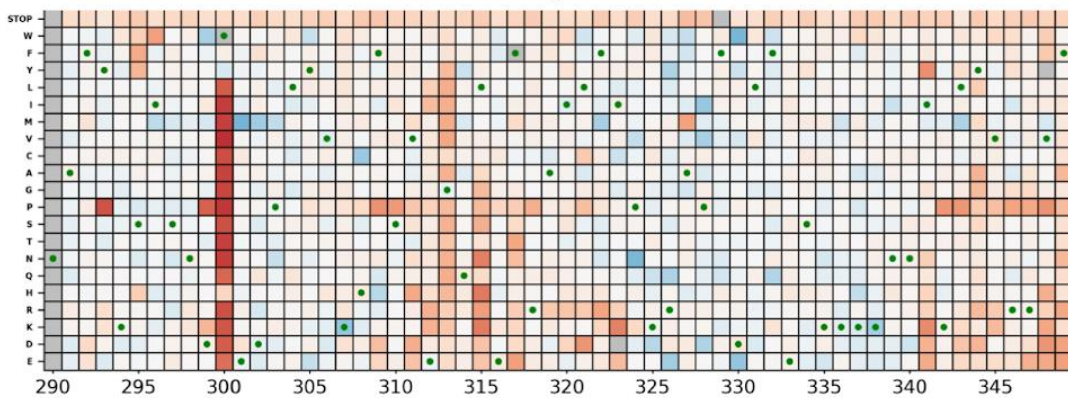
## High Salinity Replicate 2



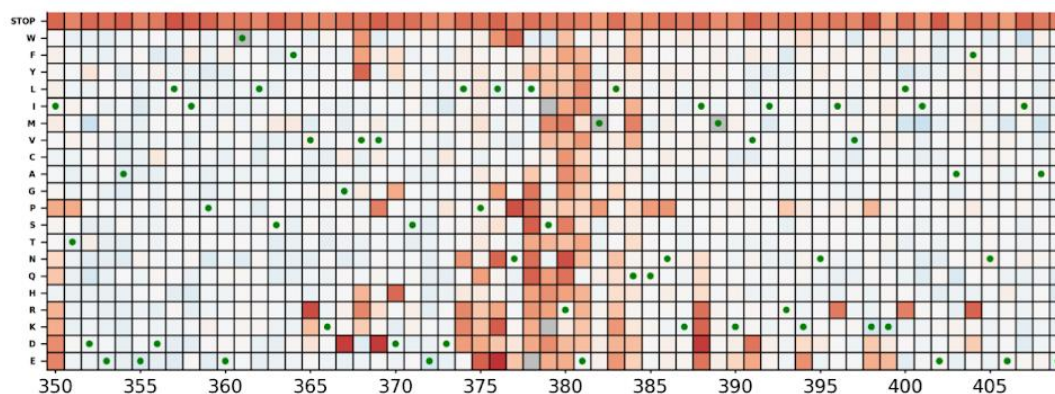
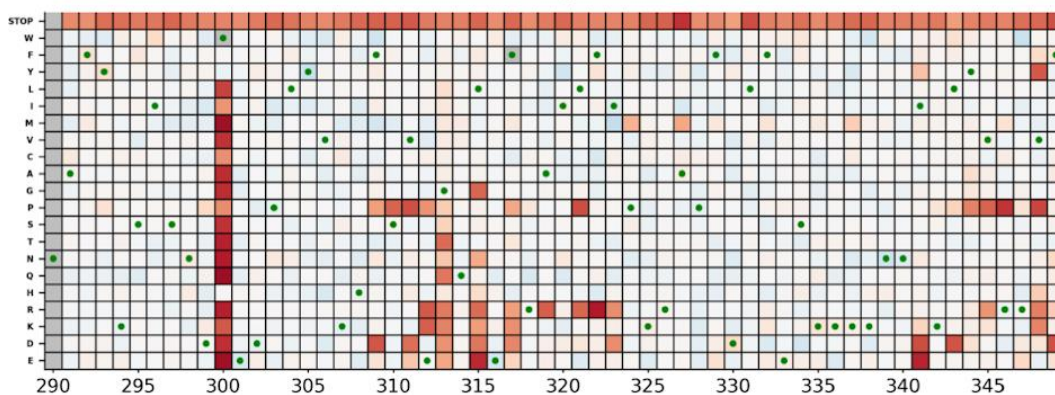
## Sorbitol Replicate 1



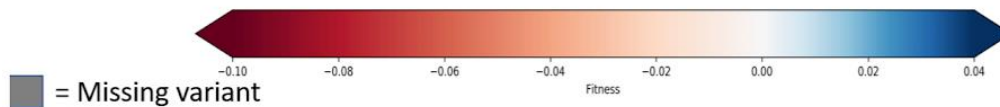
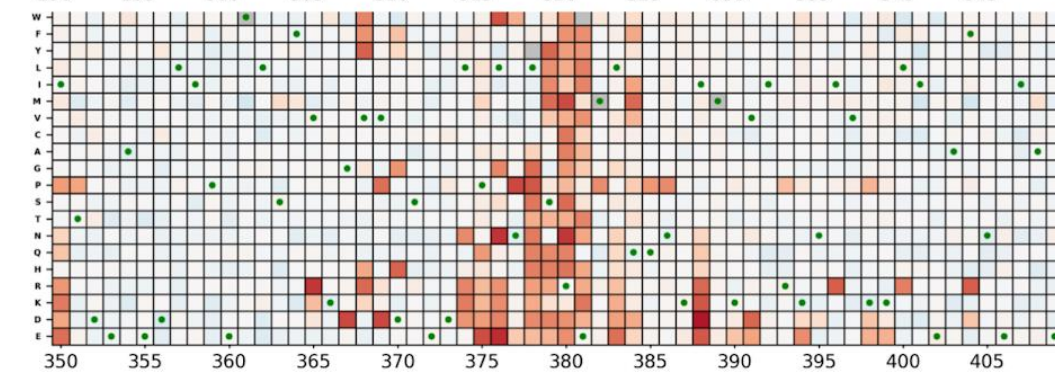
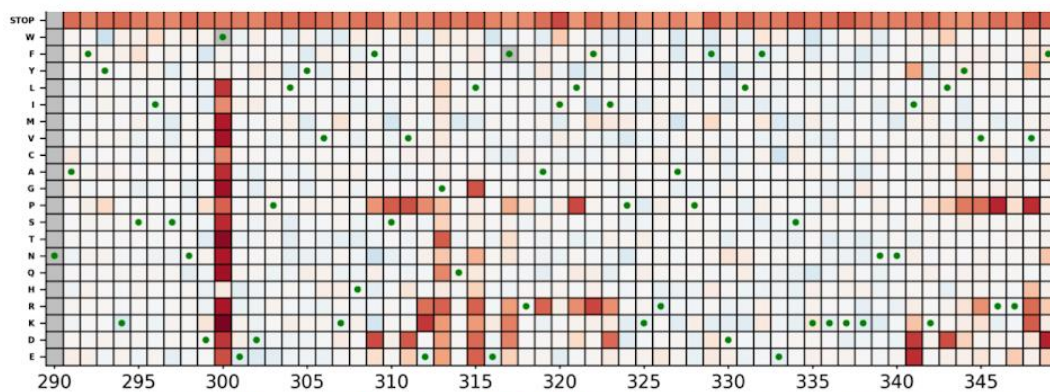
## Sorbitol Replicate 2



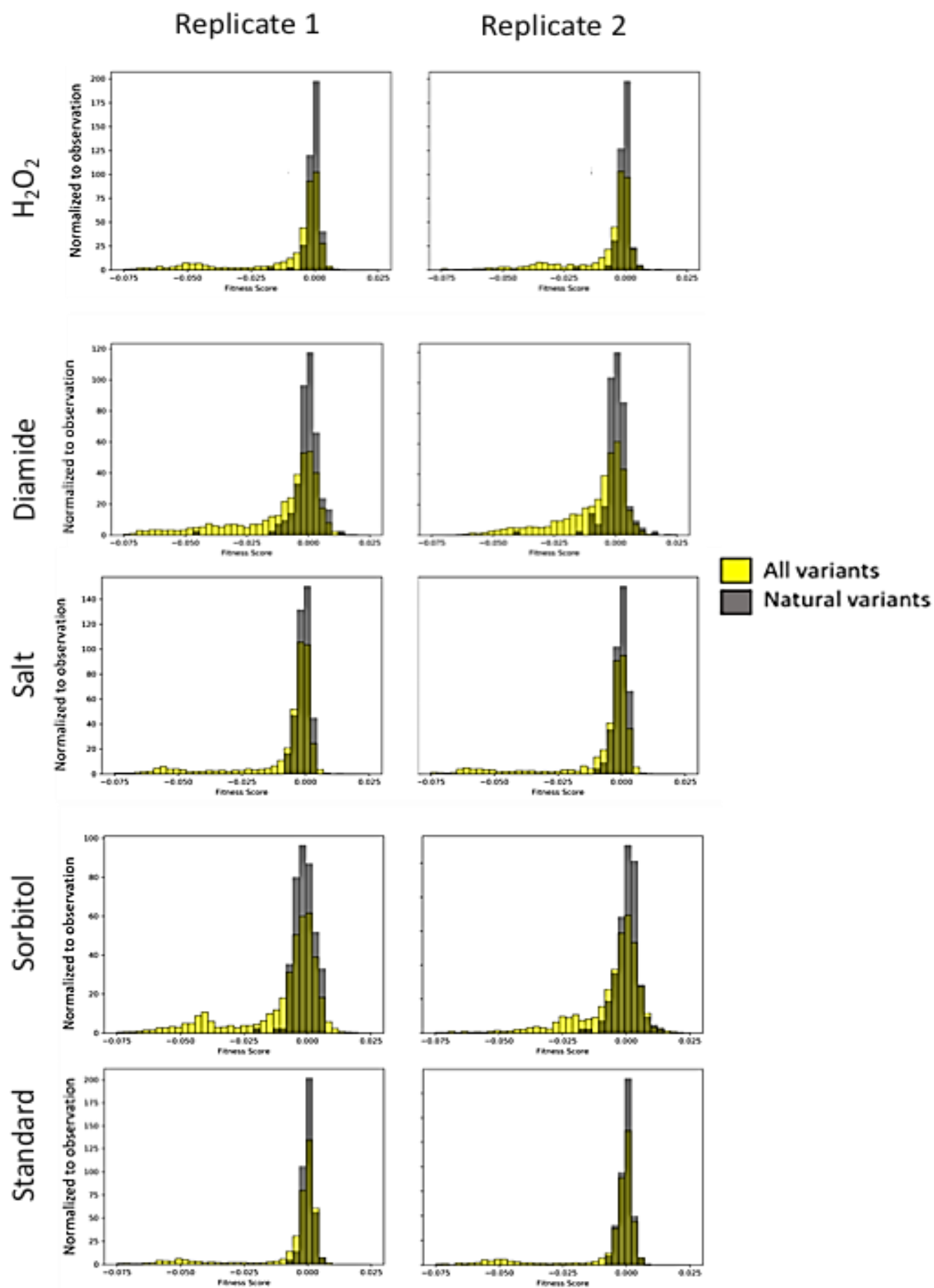
## Standard Replicate 1



## Standard Replicate 2



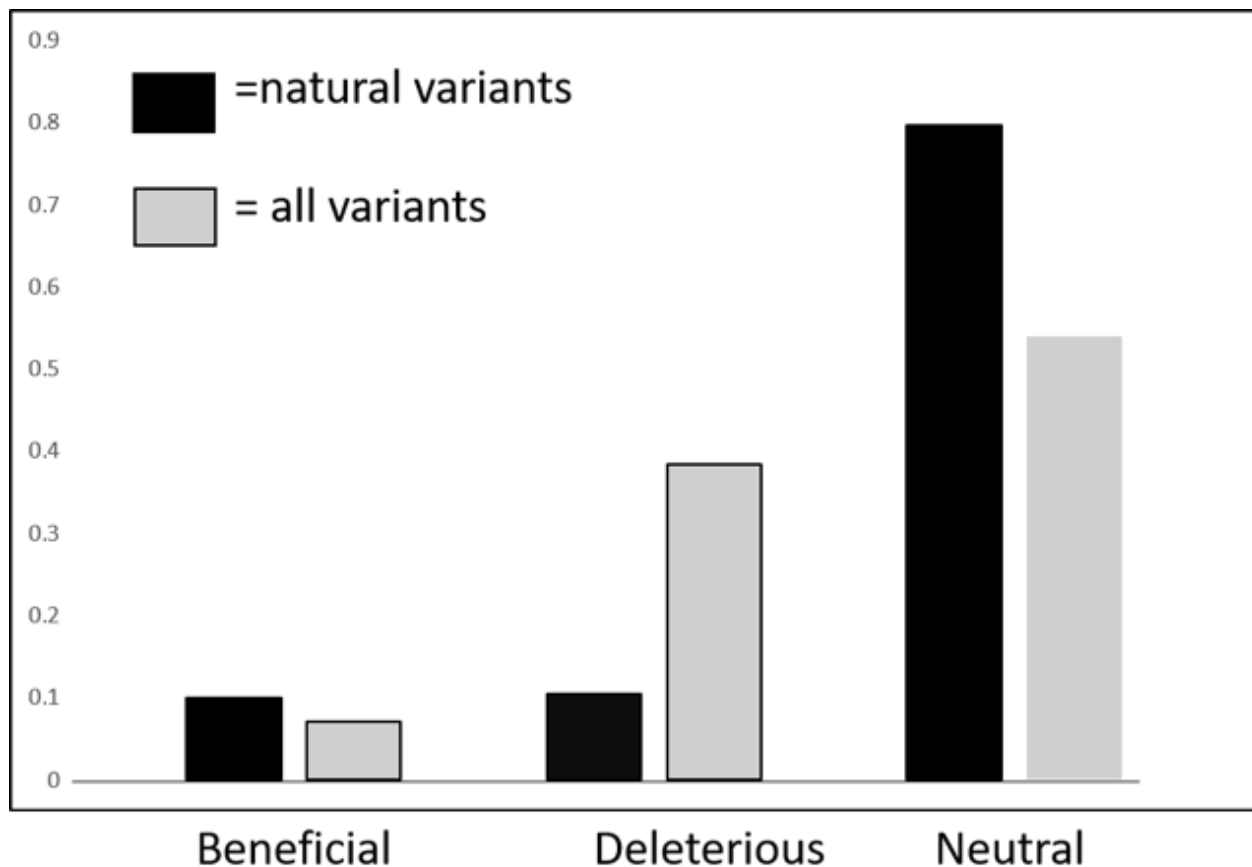
**Supplementary Figure 2.3. Heatmaps of mutant fitness effects show that the majority of mutations are wild type under all conditions, with the exception of amino acid regions ~300, 315-320, 347 and 375-390 that show mostly deleterious mutant fitness effects. Beneficial mutations were found under all conditions but in different regions of the middle domain.** (A) Diamide heatmaps under normal expression. (B) H<sub>2</sub>O<sub>2</sub> heatmaps under normal expression. (C) Salt heatmaps under and normal expression. (D) Sorbitol heatmaps under normal expression. (E) Standard heatmaps under and normal expression.





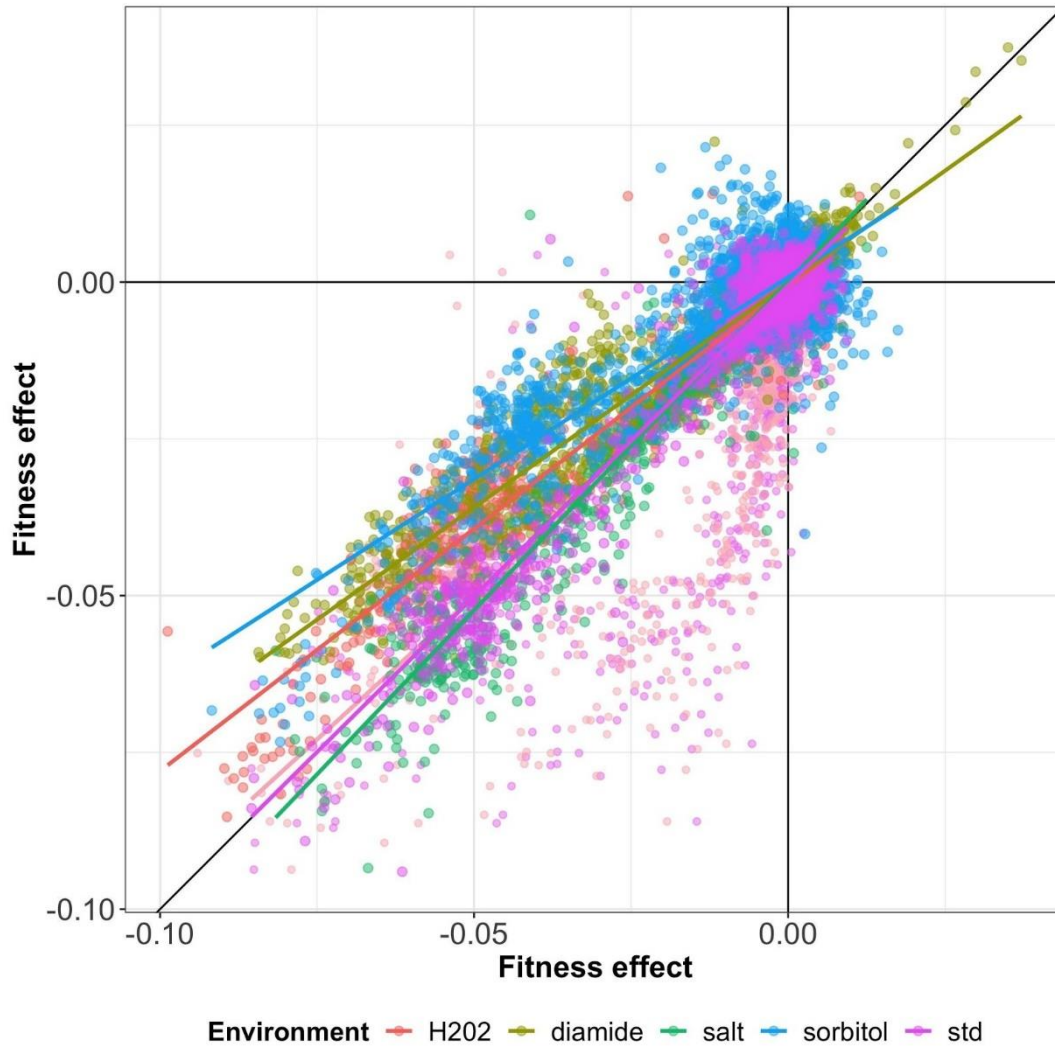
**Supplementary Figure 2.4:**

**Distribution of selection coefficients of mutations found in natural environments vs all mutations studied here.** The x-axis corresponds to the selection coefficient of the mutations and the y axis to counts of mutations normalized to the number of observations. The rows indicate different environments and the columns correspond to 2 replicates for each environment.



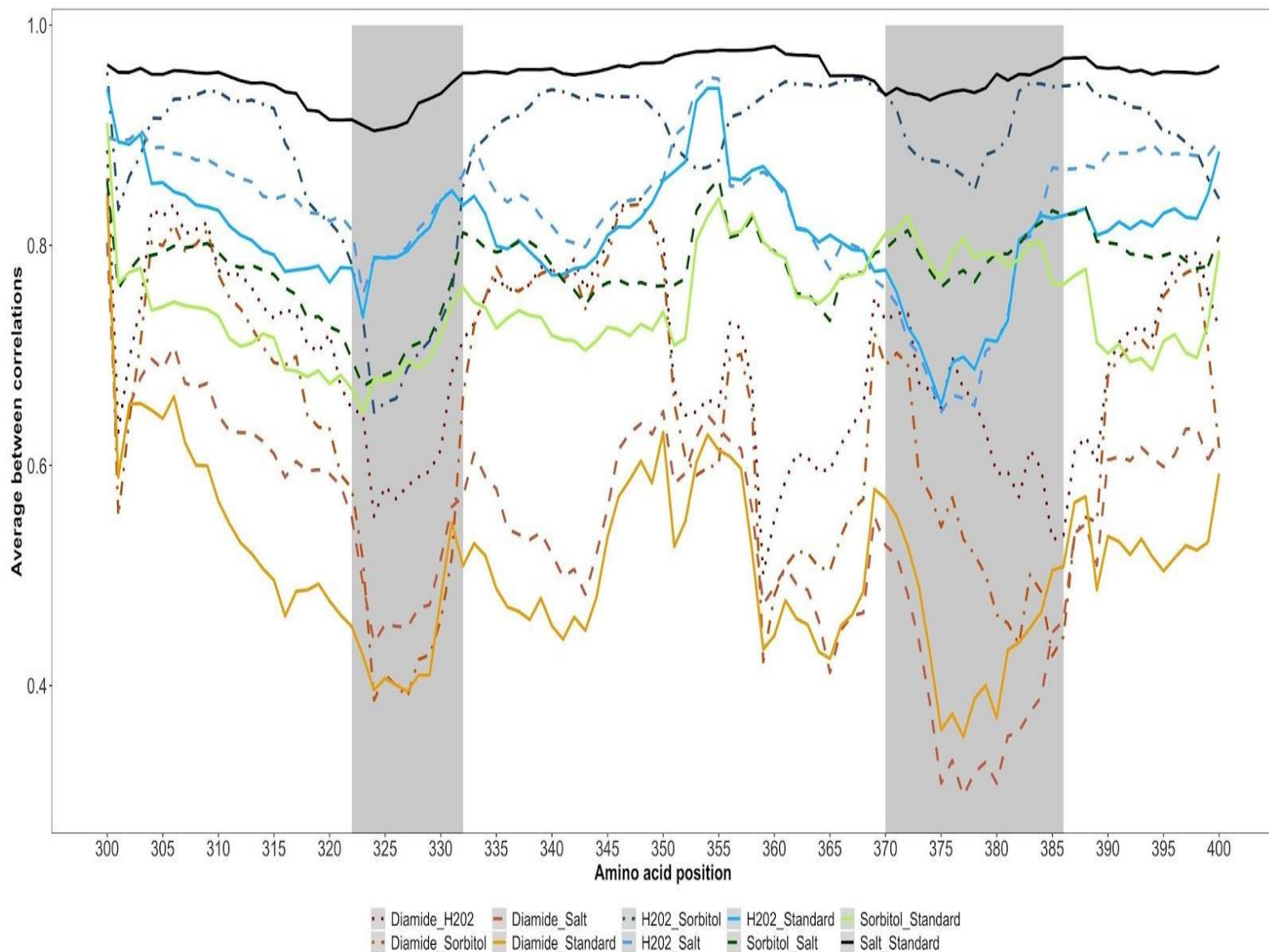
**Supplementary Figure 2.5:**

**Fraction of mutations that are beneficial, deleterious or neutral from our study and occurring in natural populations.** The x-axis corresponds to the 3 mutational categories and the y axis to the fractions of mutations.

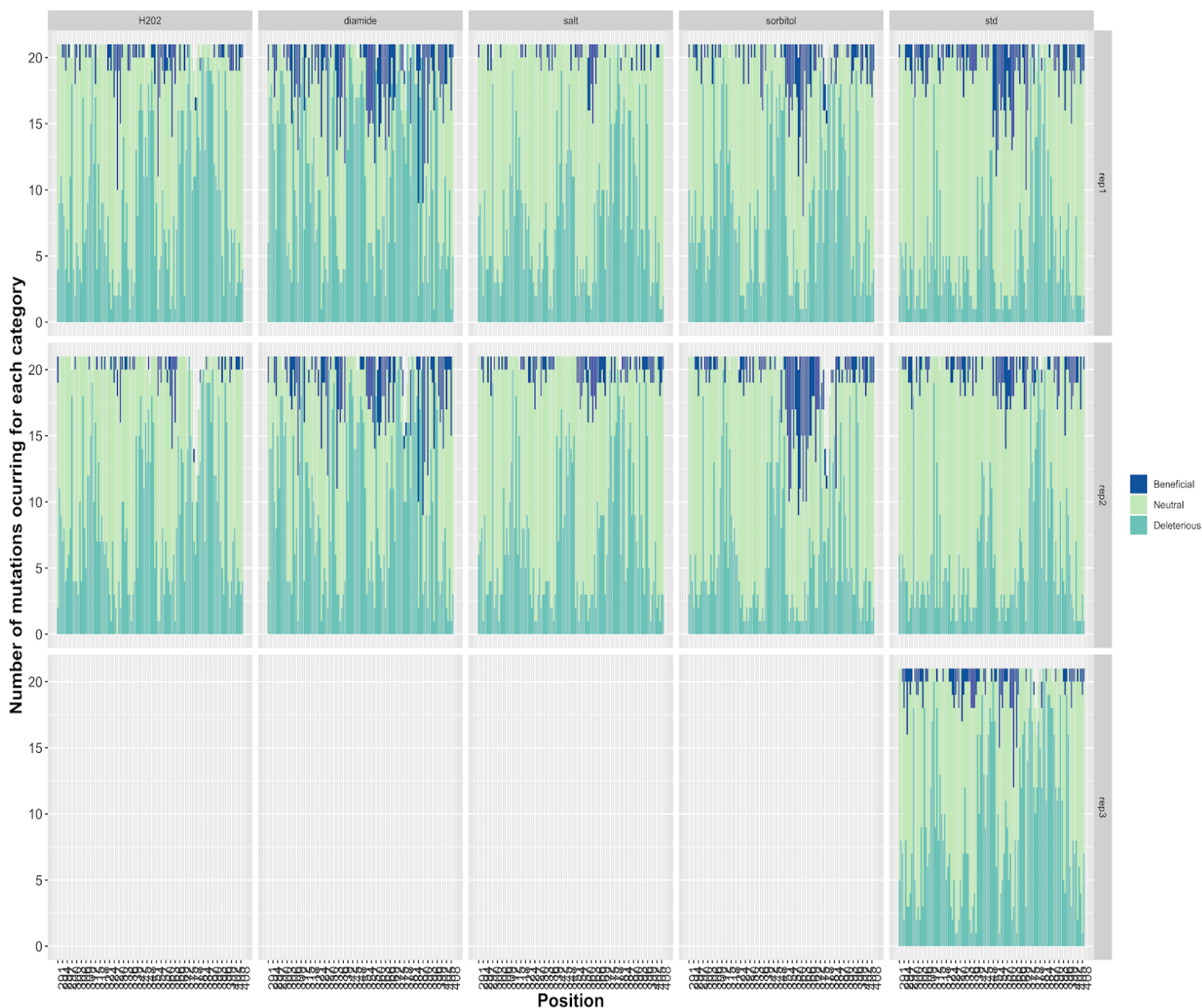


**Supplementary Figure 2.6: Correlation between replicates for each environment.**

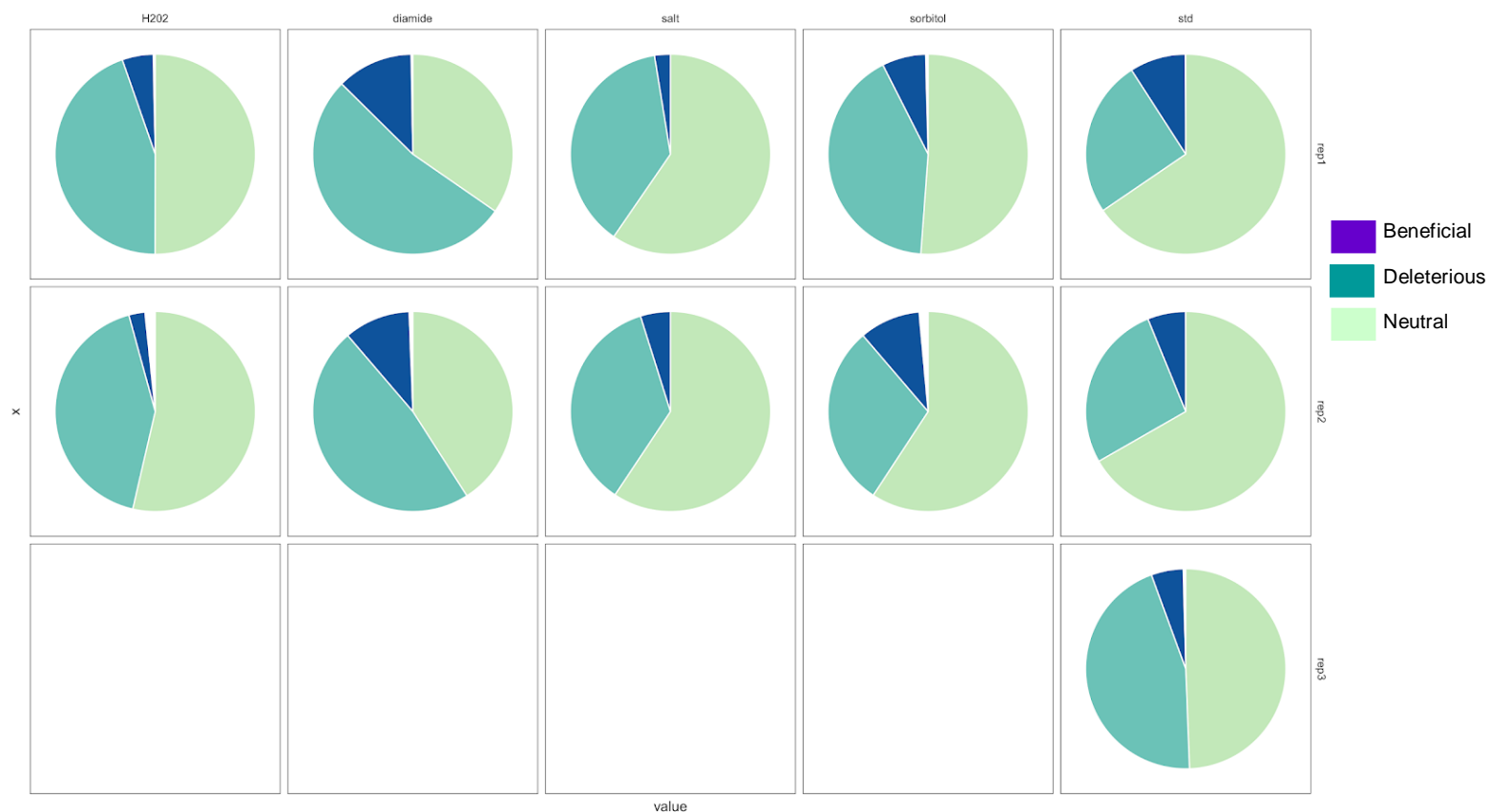
The x-axis corresponds to the fitness effect of replicate 1 (or replicate 3 for the comparison between standard environment replicate 2 vs 3) and the y axis to the fitness effect of replicate 2 (or 3 for the comparison between standard environment replicate 1 vs 3). In general we observe a very high correlation between replicates within each environment.



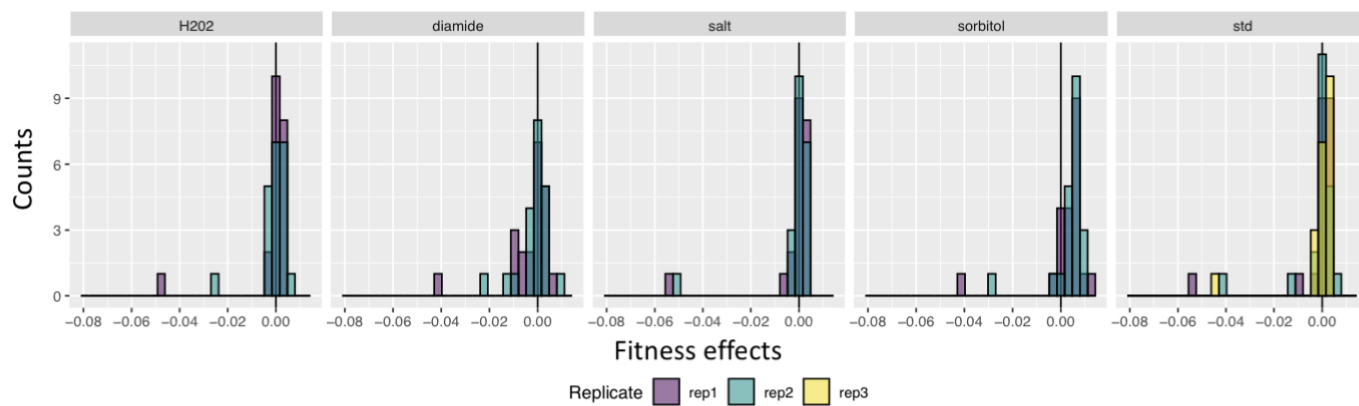
**Supplementary Figure 2.7: Correlation between environments in 10 amino acid regions.** The x-axis corresponds to amino acid positions and the y-axis to the average correlation between environments. Regions with a consistent decrease in correlations for all combinations of environments are highlighted in grey.



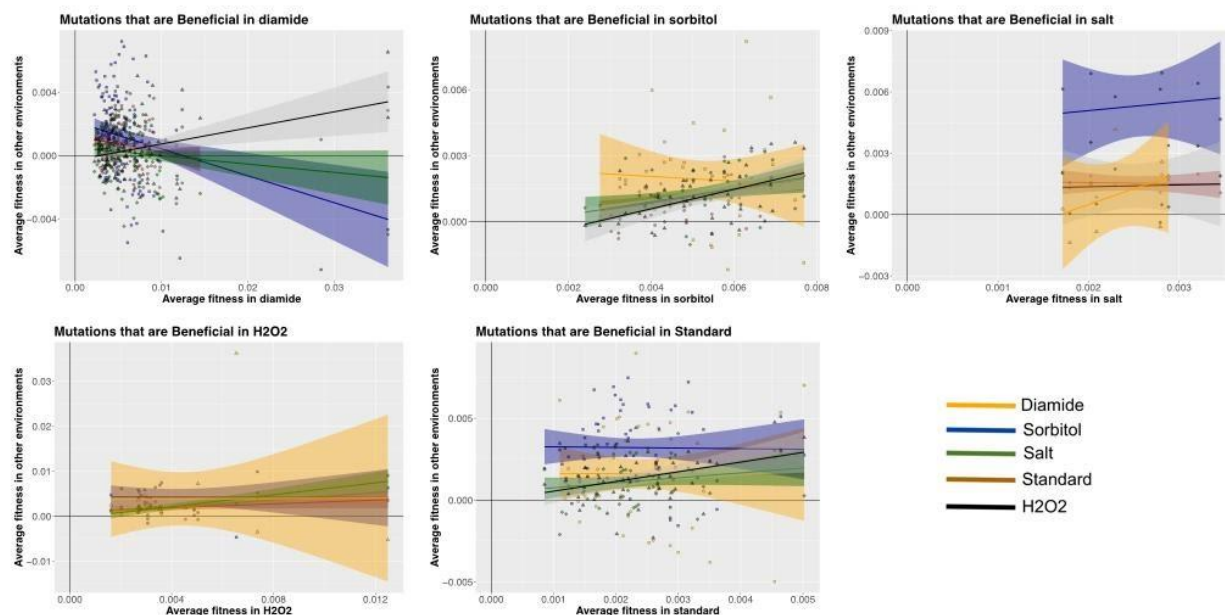
**Supplementary Figure 2. 8: Proportions of deleterious, beneficial and wild-type like mutations along the middle domain of Hsp90.** The x-axis corresponds to the amino acid positions studied and the y axis to the proportion of deleterious, beneficial and neutral mutations at each amino acid position. The rows indicate expression level and replicate, the columns the different environments.



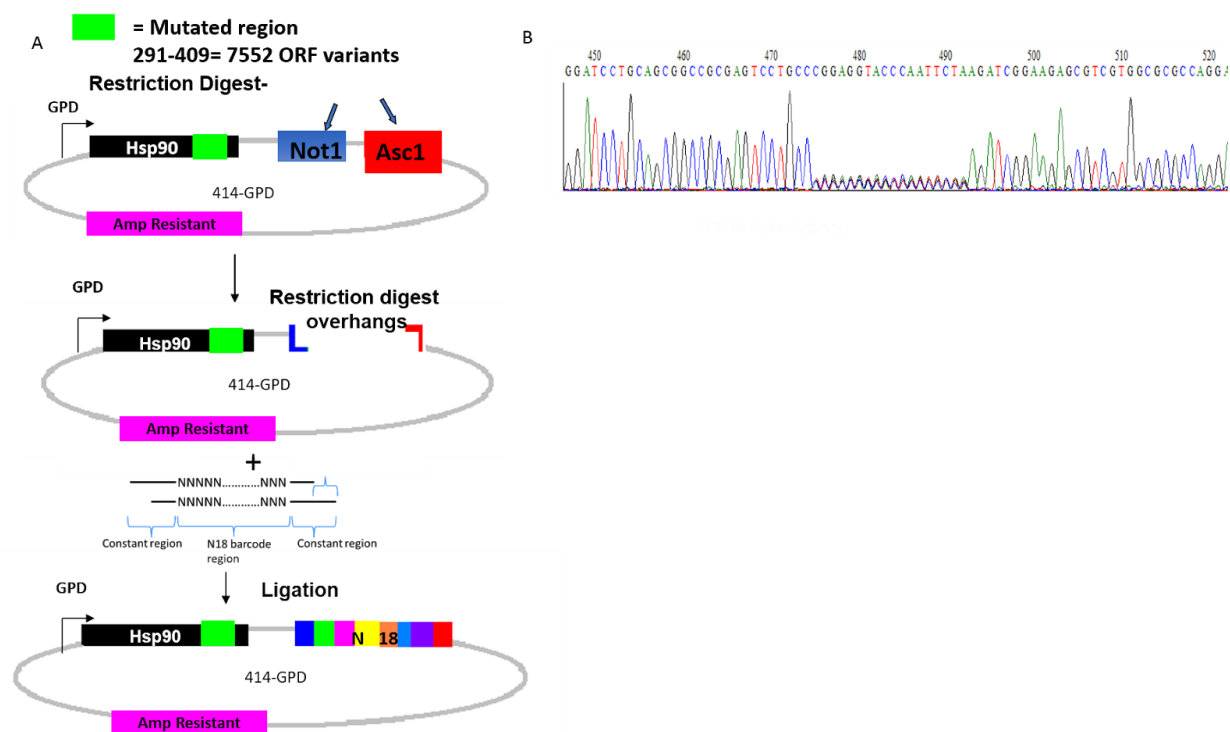
**Supplementary Figure 2.9: Summary of the proportions of beneficial, deleterious and neutral mutations in each environment/replicate.** In general, we see few beneficial mutations, except in diamide. In diamide, there is also a larger proportion of deleterious mutations in comparison with other environments.



**Supplementary Figure 2.10: Distribution of fitness effects at position 364.** Compared with the DFE of the whole middle domain (Supplementary Figure 2), there is a clear enrichment of beneficial mutations in position 364.

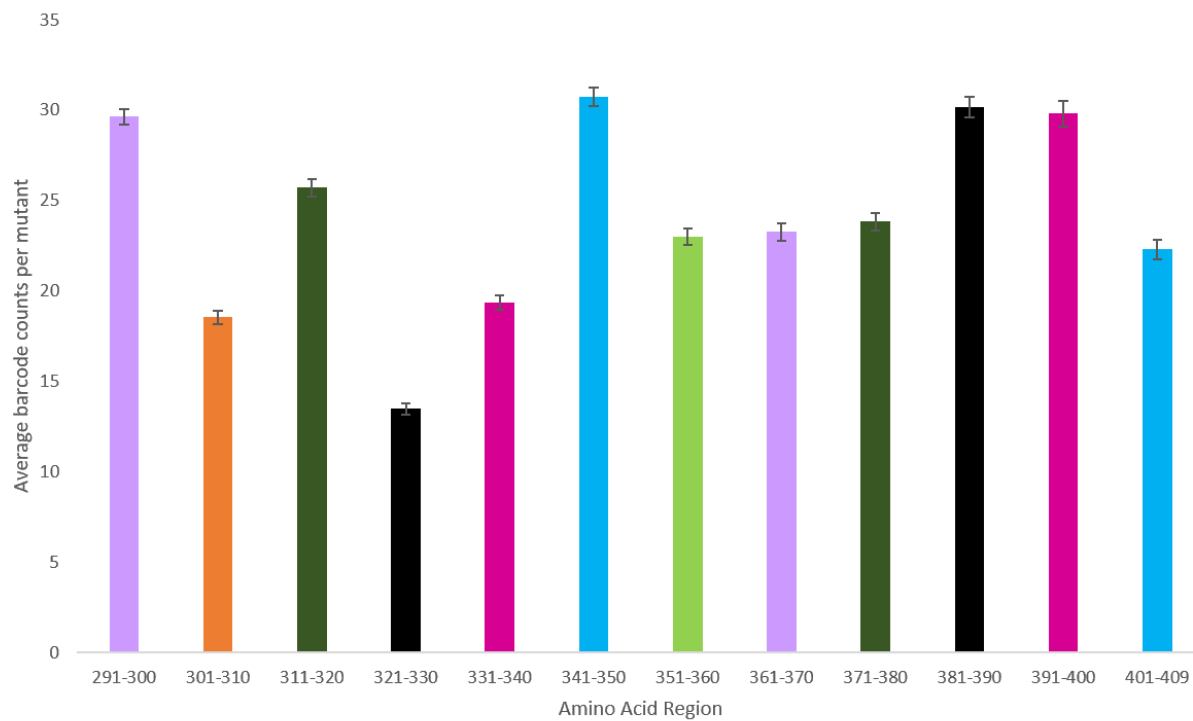


**Supplementary Figure 2.11: Cost of adaptation in different environments.** Each plot represents the average fitness effect of mutations that are beneficial in a specific environment (x-axis) and the average fitness effect in other environments (y-axis). Focal environments are indicated by the color of the lines and points. Note that the x and y scales vary between panels. We calculated the average fitness effects between the two (or three replicates) and mutations were recategorized based on whether mutations were beneficial in two replicates (beneficial mutations), beneficial in one replicate and neutral in another (beneficial-neutral), neutral in both replicates (neutral mutations), neutral in one replicate and deleterious in the other (neutral-deleterious mutations), deleterious in both replicates (deleterious) or deleterious in one replicate in beneficial in the other (beneficial-deleterious mutations).

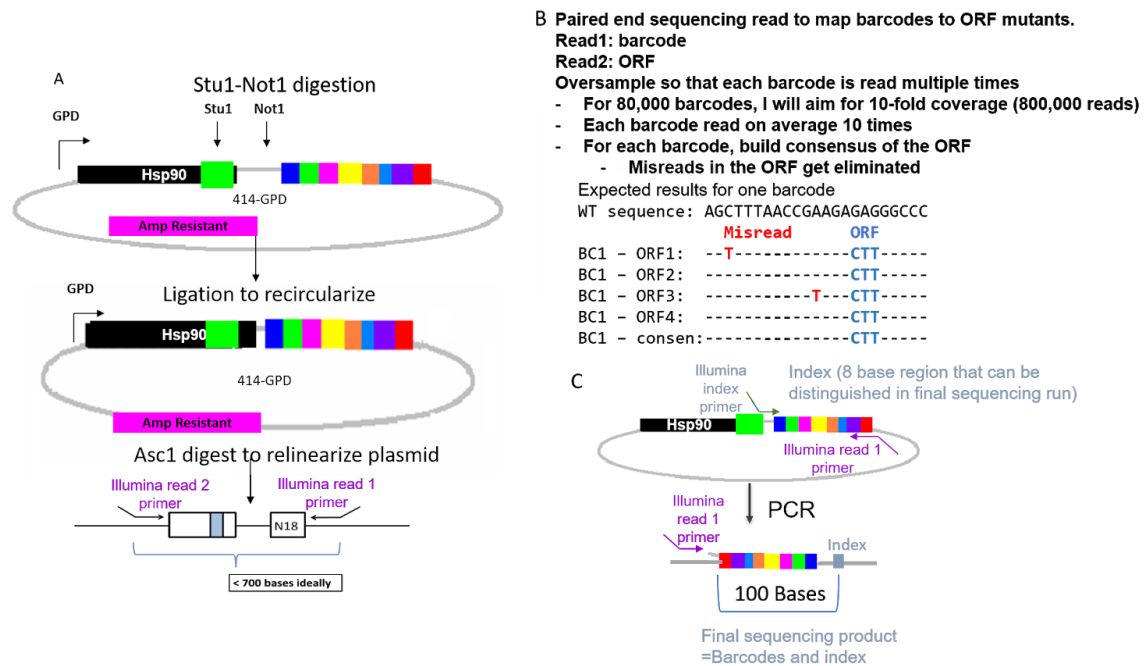


**Supplementary Figure 2.12:** Barcoding mutant library strategy. (A) Schematic of how to incorporate barcodes into plasmid libraries using restriction digests and ligation. (B) Sanger sequence TRACE result of barcoded region of the library at positions 475-493.





**Supplementary Figure 2.13:** Average number of barcodes per mutant for each 9 amino acid window shows an average of ~20 barcodes associated with each variant. The x-axis corresponds to groups of 10 amino acid regions and the y-axis to the average number of barcode counts per mutant.



**Supplementary Figure 2.14:** Strategy how to associate barcodes to mutants. A) Schematic of how to prepare barcoded mutant libraries for paired end deep sequencing. B) Schematic of how to map barcodes to open reading frame mutants. C) Schematic of how to include an index primer to distinguish each time point for final sequencing read following competition experiments.

## Supplementary Tables

### Supplementary Table 2.1

**Number of beneficial, neutral and deleterious mutations across environments, and replicates.** We observed a large proportion of wild-type like mutations in all environments.

Environment	Replicate	% Beneficial	% Neutral	% Deleterious	Total Mutations
H2O2	1	5.03 (125)	50.2 (1248)	44.77 (1113)	2486
	2	2.65 (65)	54.5 (1337)	42.85 (1051)	2453
Diamide	1	12.34 (307)	34.74 (864)	52.92 (1316)	2487
	2	10.73 (266)	41.19 (1021)	48.08 (1192)	2479
High Salinity	1	2.57 (64)	59.58 (1486)	37.85 (944)	2494
	2	4.85 (121)	59.34 (1480)	35.81 (893)	2494
Sorbitol	1	7.09 (176)	51.37 (1276)	41.55 (1032)	2484
	2	9.97 (245)	60.09 (1477)	29.94 (736)	2458
Standard	1	9.03 (225)	65.53 (1633)	25.44 (634)	2492
	2	6.14 (153)	66.75 (1664)	27.12 (676)	2493
	3	5.19 (129)	49.6 (1232)	45.21 (1123)	2484

### Supplementary Table 2.2

**Pearson correlation of fitness effects across replicates and environments.** Blue indicates stronger correlations and red weak correlations. Double lines indicate correlations between replicates of the same environment. In general the lowest correlations are seen for diamide and salt. Correlations between replicates are in general strong, except for standard replicate 3.

		Diamide		H2O2		Sorbitol		Salt		Standard		
		Rep 1	Rep 2	Rep 1	Rep 2	Rep 1	Rep 2	Rep 1	Rep 2	Rep 1	Rep 2	Rep 3
Diamide	Rep 1											
	Rep 2	0.96										
H2O2	Rep 1	0.79	0.68									
	Rep 2	0.78	0.70	0.96								
Sorbitol	Rep 1	0.76	0.66	0.95	0.94							
	Rep 2	0.69	0.62	0.88	0.90	0.87						
Salt	Rep 1	0.64	0.48	0.86	0.78	0.84	0.73					
	Rep 2	0.65	0.50	0.86	0.80	0.85	0.73	0.97				
Standard	Rep 1	0.63	0.47	0.86	0.77	0.82	0.71	0.96	0.95			
	Rep 2	0.64	0.48	0.86	0.79	0.84	0.72	0.94	0.96	0.97		
	Rep 3	0.81	0.71	0.99	0.97	0.95	0.90	0.83	0.83	0.82	0.83	

Correlation between replicates of the same environment

### Supplementary Table 2.3

#### Percentage (and number) of beneficials and their cost across environments.

Beneficials in a focal environment (rows) are deleterious in the environment indicated by the column. The diagonal of the matrix indicates the number of beneficials found in each environment/replicate. Yellow shading indicates "spurious costs of adaptation" between pairs of replicates, which is a proxy for the expected classification error. In general, this error is low (between 0.38% and 6.61%), except for Replicate 3 in the standard environment and replicate 2 from Sorbitol, where many beneficial mutations in one replicate are classified as deleterious in the other.

		Deleterious in												
		Diamide		Salt		Standard			H2O2		Sorbitol			
		Rep1	Rep2	Rep1	Rep2	Rep1	Rep2	Rep3	Rep1	Rep2	Rep1	Rep2		
Beneficial in	Diamide	Rep1	307	0 (0)	16.61 (51)	18.24 (56)	6.84 (21)	8.14 (25)	24.43 (75)	26.38 (81)	28.01 (86)	22.8 (70)	14.01 (43)	
		Rep2	0.38 (1)	266	19.17 (51)	18.05 (48)	9.02 (24)	9.02 (24)	28.95 (77)	30.83 (82)	30.45 (81)	25.19 (67)	15.79 (42)	
	Salt	Rep1	26.56 (17)	25 (16)	64	6.25 (4)	0 (0)	4.69 (3)	1.56 (1)	3.12 (2)	9.38 (6)	7.81 (5)	1.56 (1)	
		Rep2	28.93 (35)	23.14 (28)	6.61 (8)	121	4.13 (5)	0.83 (1)	9.09 (11)	11.57 (14)	11.57 (14)	6.61 (8)	0.83 (1)	
	Standard	Rep1	28.89 (65)	27.11 (61)	7.56 (17)	10.22 (23)	225	2.22 (5)	20.44 (46)	20 (45)	20.44 (46)	19.11 (43)	8.89 (20)	
		Rep2	33.99 (52)	32.68 (50)	9.15 (14)	5.88 (9)	3.92 (6)	153	24.18 (37)	24.18 (37)	19.61 (30)	12.42 (19)	5.88 (9)	
		Rep3	22.48 (29)	19.38 (25)	4.65 (6)	4.65 (6)	2.33 (3)	3.88 (5)	129	0 (0)	10.08 (13)	2.33 (3)	0 (0)	
	H2O2	Rep1	24 (30)	17.6 (22)	4.8 (6)	3.2 (4)	2.4 (3)	4 (5)	0.8 (1)	125	4.8 (6)	6.4 (8)	1.6 (2)	
		Rep2	21.54 (14)	16.92 (11)	4.62 (3)	1.54 (1)	6.15 (4)	1.54 (1)	4.62 (3)	0 (0)	65	3.08 (2)	1.54 (1)	
	Sorbitol	Rep1	27.84 (49)	23.3 (41)	5.68 (10)	2.27 (4)	2.27 (4)	2.84 (5)	5.11 (9)	7.39 (13)	10.23 (18)	176	3.98 (7)	
		Rep2	34.29 (84)	29.8 (73)	7.35 (18)	3.67 (9)	4.9 (12)	4.08 (10)	12.24 (30)	16.73 (41)	13.06 (32)	15.1 (37)	245	
					Percentage of beneficials that have deleterious effect between replicates									
					Number of beneficials in each environment									

### **Chapter III: Comprehensive fitness maps of Hsp90 show widespread environmental dependence**

This work has been published previously as *Julia M. Flynn<sup>1</sup>, Ammeret Rossouw<sup>1</sup>, Pamela A. Cote-Hammarlof<sup>1</sup>, Ines Fragata<sup>2</sup>, David Mavor<sup>1</sup>, Carl Hollins III<sup>1</sup>, Claudia Bank<sup>2</sup>, and Daniel N.A. Bolon<sup>\*</sup>. Comprehensive fitness maps of Hsp90 show widespread environmental dependence. *Elife* 9: e53810. Published 2020 Mar 4. doi:10.7554/eLife.53810*

This was a collaborative effort. I, Pamela Cote-Hammarlof contributed to the conceptualization and experimental design for this study. I transformed yeast with Hsp90 mutant libraries and performed growth competitions for one of the experimental replicates. I isolated DNA from these samples and prepared them for deep sequencing. I assisted in initial deep sequencing analysis of samples. I contributed to reviewing and editing the manuscript.

**Abstract**

Gene-environment interactions have long been theorized to influence molecular evolution. However, the environmental dependence of most mutations remains unknown. Using deep mutational scanning, we engineered yeast with all 44,604 single codon changes encoding 14,160 amino acid variants in Hsp90 and quantified growth effects under standard conditions and under five stress conditions. To our knowledge, these are the largest determined comprehensive fitness maps of point mutants. The growth of many variants differed between conditions, indicating that environment can have a large impact on Hsp90 evolution. Multiple variants provided growth advantages under individual conditions; however, these variants tended to exhibit growth defects in other environments. The diversity of Hsp90 sequences observed in extant eukaryotes preferentially contains variants that supported robust growth under all tested conditions. Rather than favoring substitutions in individual conditions, the long-term selective pressure on Hsp90 may have been that of fluctuating environments, leading to robustness under a variety of conditions.

## Introduction

The role of environment has been contemplated in theories of evolution for over a hundred years (Darwin and Wallace 1858; Darwin 1859; Wright 1932), yet molecular level analyses of how environment impacts the evolution of gene sequences remain experimentally under-explored. Depending on environmental conditions, mutations can be categorized into three classes: strongly deleterious mutations that are purged from populations by purifying selection, nearly-neutral mutations that are governed by stochastic processes, and beneficial mutations that tend to provide a selective advantage (Ohta 1973). It has long been clear that environmental conditions can alter the fitness effects of mutations (Tutt 1896). However, examining how environmental conditions impact any of the three classes of mutations is challenging. Measurable properties of nearly-neutral and deleterious mutations in natural populations are impacted by both demography and selection (Ohta 1973), which are difficult to disentangle. In addition, many traits are complex, making it challenging to identify all contributing genetic variations (McCarthy et al. 2008). For these and other reasons, we do not have a detailed understanding of how environmental conditions impact the evolution of most gene sequences.

Mutational scanning approaches (Fowler et al. 2010) provide novel opportunities to examine fitness effects of the same mutations under different laboratory conditions (Boucher et al. 2014; Boucher et al. 2016; Canale et al. 2018; Kemble et al. 2019). The EMPIRIC (Exceedingly Meticulous and Parallel Investigation of Randomized Individual Codons) approach that we previously developed (Hietpas et al. 2011) is particularly well suited to address questions regarding the environmental impact of mutational effects for



three reasons: it quantifies growth rates that are a direct measure of experimental fitness, all point mutations are engineered providing comprehensive maps of growth effects, and all the variants can be tracked in the same flask while experiencing identical growth conditions. We have previously used the EMPIRIC approach to investigate how protein fitness maps of ubiquitin vary in different environmental conditions (Mavor et al. 2016). The analysis of ubiquitin fitness maps revealed that stress environments can exacerbate the fitness defects of mutations. However, the small size of ubiquitin and the near absence of natural variation in ubiquitin sequences (only three amino acid differences between yeast and human) hindered investigation of the properties underlying historically observed substitutions.

Mutational scanning approaches have emerged as a robust method to analyze relationships between gene sequence and function, including aspects of environmental-dependent selection pressure. Multiple studies have investigated resistance mutations that enhance growth in drug or antibody environments (Firnberg et al. 2014; Stiffler et al. 2015; Doud and Bloom 2016; Jiang et al. 2016; Dingens et al. 2018). Most of these studies have focused on interpreting adaptation in the light of protein structure. Of note, Dandage, Chakraborty and colleagues explored how environmental perturbations to protein folding influenced tolerance of mutations in the 178 amino acid gentamicin-resistant gene in bacteria (Dandage et al. 2017). However, the question of how environmental variation shapes the selection pressure on gene sequences has not been well studied.

Here, we report comprehensive experimental fitness maps of Heat Shock Protein 90 (Hsp90) under multiple stress conditions and compare our experimental results with the historical record of hundreds of Hsp90 substitutions accrued during its billion years of evolution in eukaryotes. Hsp90 encodes a 709 amino acid protein and to our knowledge it is the largest gene for which a comprehensive protein fitness map has been determined. Hsp90 is an essential and highly abundant molecular chaperone which is induced by a wide variety of environmental stresses (Lindquist 1981; Gasch et al. 2000). Hsp90 assists cells in responding to these stressful conditions by facilitating the folding and activation of client proteins through a series of ATP-dependent conformational changes mediated by co-chaperones (Krukenberg et al. 2011). These clients are primarily signal transduction proteins, highly enriched in kinases and transcription factors (Taipale et al. 2012). Through its clients, Hsp90 activity is linked to virtually every cellular process.

Hsp90 can facilitate the emergence and evolution of new traits in response to stress conditions, including drug resistance in fungi (Cowen and Lindquist 2005), gross morphology in flies (Rutherford and Lindquist 1998) and plants (Queitsch et al. 2002), and vision loss in cave fish (Rohner et al. 2013). In non-stress conditions, an abundance of Hsp90 promotes standing variation by masking the phenotypic effects of destabilizing mutations in clients. Stressful conditions that tax Hsp90 capacity can then manifest in phenotypic diversity that can contribute to adaptation. Because of the biochemical and evolutionary links between Hsp90 and stress, we hypothesized that environmental stress would result in altered fitness maps.

The conditions in natural environments often fluctuate, and all organisms contain stress response systems that aid in acclimation to new conditions. The conditions experienced by different populations can vary tremendously depending on the niches that they inhabit, providing the potential for distinct selective pressures on Hsp90. Previous studies of a nine amino acid loop in Hsp90 identified multiple amino acid changes that increased the growth rate of yeast in elevated salinity (Hietpas et al. 2013), demonstrating the potential for environmental-dependent beneficial mutations in Hsp90. However, the sequence of Hsp90 is strongly conserved in eukaryotes (57% amino acid identity from yeast to human), indicating consistent strong purifying selection.

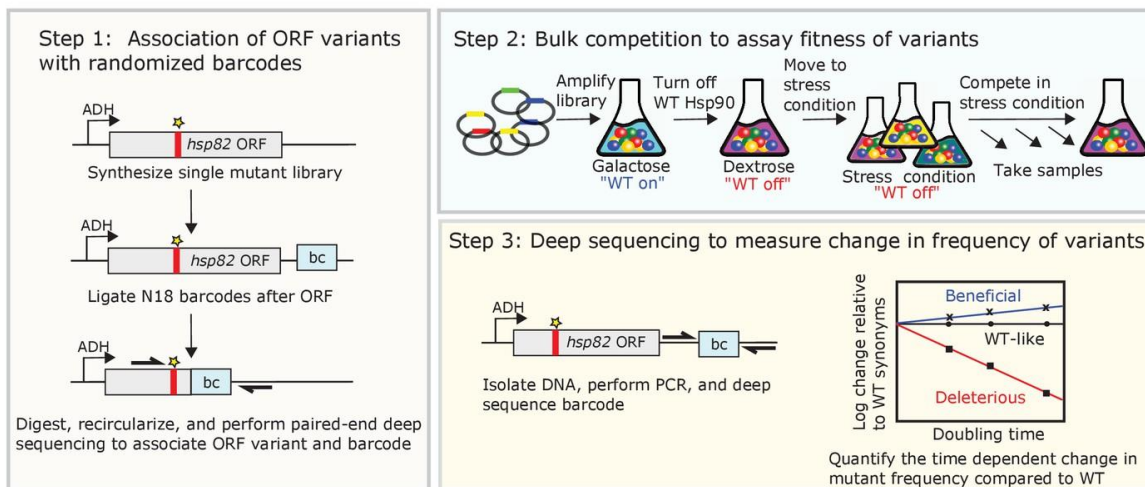
To investigate the potential influence of the environment on Hsp90 evolution, we quantified fitness maps in six different conditions. While proximity to ATP is the dominant functional constraint in standard conditions, the influence of client and co-chaperone interactions on growth rate dramatically increases under stress conditions. Increased selection pressure from heat and diamide stresses led to a greater number of beneficial variants compared to standard conditions. The observed beneficial variants were enriched at functional hotspots in Hsp90. However, the natural variants of Hsp90 tend to support efficient growth in all environments tested, indicating selection for robustness to diverse stress conditions in the natural evolution of Hsp90.

## **Results**

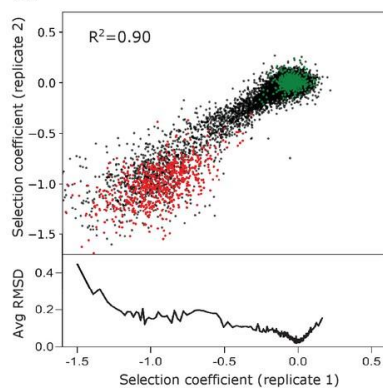
We developed a powerful experimental system to analyze the growth rate supported by all possible Hsp90 point mutations under distinct growth conditions. Bulk

competitions of yeast with a deep sequencing readout enabled the simultaneous quantification of 98% of possible amino acid changes (Figure 3.1A). The single point mutant library was engineered by incorporating a single degenerate codon (NNN) into an otherwise wild-type Hsp90 sequence as previously described (Hietpas et al. 2012). To provide a sensitive readout of changes in Hsp90 function, we transferred the library to a plasmid under a constitutive low-expression level ADH promoter that reduced Hsp90 protein levels to near-critical levels (Jiang et al. 2013). To efficiently track all variants in a single competition flask so that all variants experience identical conditions, we updated our previously developed EMPIRIC approach to include a barcoding strategy (Hietpas et al. 2012). As described in the Materials and methods, this barcode strategy enabled us to track mutations across a large gene using a short sequencing readout

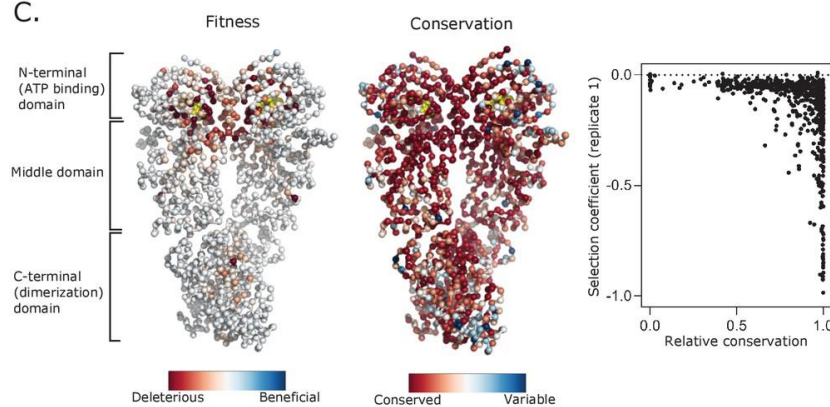
A.



B.



C.



### Figure 3.1: Approach to determine protein fitness maps of Hsp90.

(A) Barcoded competition strategy to analyze the growth effects of all single codon variants of Hsp90 in a single bulk culture. *Hsp82* is the stress-inducible gene that encodes for Hsp90 (B) Measurements of selection coefficients of amino acid variants are reproducible in replicate growth competitions (see Figure 3.1—source data). Wild-type amino acids are shown in green and stop codons are shown in red. The bottom panel shows the Root-mean-square deviation (RMSD) averaged for a running window of 40 data points. (C) Average selection coefficients at each position in standard conditions mapped onto a homodimeric structure of Hsp90 (PDB 2cg9, (Ali et al. 2006)) and compared to patterns of evolutionary conservation (see Figure 3.1-source data 2). ATP is shown in yellow. The graph on the right compares relative conservation at each position of Hsp90 to the average selection coefficient at that position.

**Figure 3.1—source data 1**

**Sequencing counts and selection coefficients for each individual amino acid change across amino acids 2–709 of Hsp90 in both replicates of standard conditions.**

<https://cdn.elifesciences.org/articles/53810/elifesciences-53810-fig1-data1-v2.xlsx>

[Download elifesciences-53810-fig1-data1-v2.xlsx](#)

**Figure 3.1—source data 2**

**Average selection coefficient (excluding stops) at each position of Hsp90 in Standard replicate 1.**

<https://cdn.elifesciences.org/articles/53810/elifesciences-53810-fig1-data2-v2.xlsx>

[Download elifesciences-53810-fig1-data2-v2.xlsx](#)

We transformed the plasmid library of comprehensive Hsp90 point mutations into a conditional yeast strain where we could turn selection of the library on or off. We used a yeast Hsp90-shutoff strain in which both paralogs of Hsp90 (*hsc82* and *hsp82*) are deleted and a copy of *hsp82* with expression under strict regulation of a galactose-inducible promoter is integrated into the chromosome (Jiang et al. 2013). The mutant libraries were amplified in the absence of selection on the mutant variant by growing the transformed yeast in galactose media that expresses the wild-type chromosomal copy of *hsp82*. We switched the yeast to dextrose media to shut off the expression of wild-type Hsp90, allowing the mutagenized variants to be the sole source of Hsp90 protein in the cell, and then split the culture into six different environmental conditions. We extracted samples from each condition at multiple time points and used Illumina sequencing to estimate the frequency of each Hsp90 variant over time. We assessed the selection coefficient of each Hsp90 variant from the change in frequency relative to wild-type Hsp90 using a previously developed Bayesian Markov Chain Monte Carlo

(MCMC) method (Bank et al. 2014; Fragata et al. 2018), where 0 represents wild-type and -1 represents null alleles (Figure 3.1- source data 1).

To analyze reproducibility of the growth competition, we performed a technical replicate under standard conditions. We used a batch of the same transformed cells that we had frozen and stored such that the repeat bulk competition experiments and sequencing were performed independently. Selection coefficients between replicates were strongly correlated ( $R^2 = 0.90$ ), and indicated that we could clearly distinguish between selection coefficients for members of the library containing silent mutations that do not change the amino acid sequence (wild-type synonyms) and those containing stop codons (Figure 3.1B, Supplementary Figure 3.1). For the second replicate we noted a small fitness defect ( $s \sim -0.2$ ) for wild-type synonyms at positions 679–709 relative to other positions (Supplementary Figure 3.1). We did not see this behavior in any other condition or replicate tested and do not understand its source. The selection coefficients in this study under standard conditions also correlated strongly ( $R^2 = 0.87$ ) with estimates of the Hsp90 N-domain in a previous study (Mishra et al. 2016); (Supplementary Figure 3.2), indicating that biological replicates also show high reproducibility. Of note, variants with strongly deleterious effects exhibited the greatest variation between replicates, consistent with the noise inherent in estimating the frequency of rapidly depleting variants (Figure 3.1B). The stop codons were already partially depleted from the cells at the 0 time point, likely contributing to their variation between replicates (Supplementary Figure 3.3A). In accordance with this, there was a higher variation in selection coefficients between replicates for stop codons with the lowest initial reads (Supplementary Figure 3.3B). Stop codon fitness was similar for all

three stop codons (Supplementary Figure 3.3C) and at positions across Hsp90 with exception of the last 32 positions that have previously been shown to be dispensable for its viability (Louvion et al. 1996); (Supplementary Figure 3.1). A heatmap representation of all the selection coefficients determined in standard conditions in replicate one is shown in (Supplementary Figure 3.4).

The large number of signaling pathways that depend on Hsp90 (Taipale et al. 2012) and its strong sequence conservation suggest that many mutations of Hsp90 may decrease fitness. However, most variants of Hsp90 had wild-type-like fitness in the competition experiment in standard conditions (Figure 3.1C, Supplementary Figure 3.4). All possible mutations (excluding stops) were compatible with function at 425 positions. Only 17 positions had low mutational tolerance to the extent that 15 or more substitutions caused null-like growth defects (R32, E33, N37, D40, D79, G81, G94, I96, A97, S99, G118, G121, G123, Y125, F156, W300, and R380). All these positions except for W300 are in contact with ATP or mediate ATP-dependent conformational changes in the N-domain of Hsp90. In fact, the average selection coefficient at different positions (a measure of mutational sensitivity) in standard growth conditions correlates ( $R^2 = 0.49$ ) with distance from ATP (Supplementary Figure 3.5). While W300 does not contact ATP, it transmits information from client binding to long range conformational changes of Hsp90 that are driven by ATP hydrolysis (Rohl et al. 2013). Our results indicate that ATP binding and the conformational changes driven by ATP hydrolysis impose dominant physical constraints in Hsp90 under standard laboratory conditions.



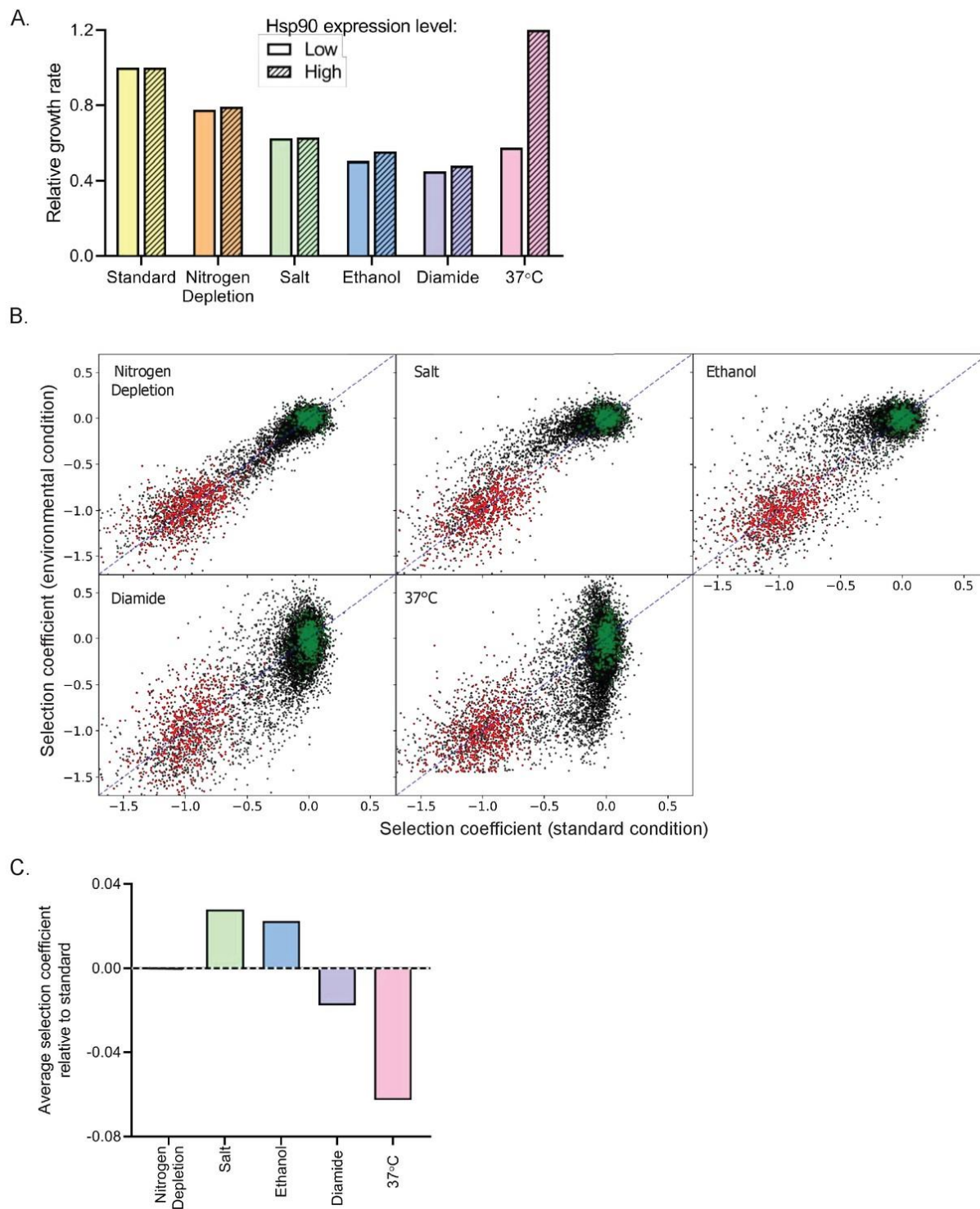
At first sight, the observation that most mutations are compatible with robust growth in standard conditions is at odds with the fact that the Hsp90 sequence is strongly conserved across large evolutionary distances (Figure 3.1C). One potential reason for this discrepancy could be that the strength of purifying selection in large natural populations over long evolutionary time-scales is more stringent than can be measured in the laboratory. In other words, experimentally unmeasurable fitness defects could be subject to purifying selection in nature. In addition, the range of environmental conditions that yeast experience in natural settings may not be reflected by standard laboratory growth conditions. To investigate the impact of environmental conditions on mutational effects in Hsp90, we measured the growth rate of Hsp90 variants under five additional stress conditions.

### **Impact of Stress Conditions on Mutational Sensitivity of Hsp90**

We measured the fitness of Hsp90 variants in conditions of nitrogen depletion (ND) (0.0125% ammonium sulfate), hyper-osmotic shock (0.8 M NaCl), ethanol stress (7.5% ethanol), the sulfhydryl-oxidizing agent diamide (0.85 mM), and temperature shock (37°C). All these stresses are known to elicit a common shared environmental stress response characterized by altered expression of ~900 genes as well as having specific responses unique to each stress (Gasch et al. 2000). Genes encoding heat-shock proteins, including Hsp90, are transiently upregulated in all these stresses except elevated salinity (Piper 1995; Gasch et al. 2000).

One way to characterize stress conditions is to measure the extent to which they slow down growth. For our experiments, each of the environmental stresses were

selected to partially decrease the growth rate. Consistently, all stresses reduced the growth rate of the parental strain within a two-fold range, with depletion of nitrogen levels causing the smallest reduction in growth rate and diamide causing the greatest reduction (Figure 3.2A). To investigate how critical Hsp90 is for growth in each condition, we measured growth rates of yeast with either normal or more than 10-fold reduced (Jiang et al. 2013) levels of Hsp90 protein (Figure 3.2A). Under standard conditions, the normal level of Hsp90 protein can be dramatically reduced without major impacts on growth rate, consistent with previous findings (Picard et al. 1990; Jiang et al. 2013).



**Figure 3.2: Impact of environmental stresses on yeast growth rates and selection on Hsp90 sequence. (A)** Growth rate of yeast with normal and reduced expression of

Hsp90 protein in standard and stress conditions based on individual growth curves. Growth rates are normalized to growth in standard conditions with reduced Hsp90 expression. **(B)** Selection coefficients of all Hsp90 amino acid variants in stress conditions compared to standard conditions (see Figure 3.2—source data 1 ). Wild-type synonyms are shown in green and stop codons are shown in red. Selection coefficients were scaled to null ( $s = -1$ ) for the average stop codon and neutral ( $s = 0$ ) for the average wild type. The diagonal is indicated by the blue dashed line. **(C)** The average selection coefficient of all mutations relative to standard conditions, a metric of the strength of selection acting on Hsp90 sequence, in each stress condition.

### **Figure 3. 2—source data 1**

**Sequencing counts and selection coefficients for each individual amino acid change across amino acids 2–709 of Hsp90 in Nitrogen Depletion, Salt, Ethanol, Diamide and 37°C.**

<https://cdn.elifesciences.org/articles/53810/elifesciences-53810-fig2-data1-v2.xlsx>

[Download elifesciences-53810-fig2-data1-v2.xlsx](#)

We anticipated that Hsp90 would be required at increased levels for robust experimental growth in diamide, nitrogen starvation, ethanol, and high temperature (Gasch et al. 2000) based on the concept that cells increase expression level of genes in conditions where those gene products are needed at higher concentration. Consistent with this concept, reduced Hsp90 levels cause a marked decrease in growth rate at 37°C. However, Hsp90 protein levels had smaller impacts on growth rates under the other stress conditions, indicating that reliance on overall Hsp90 function does not increase dramatically in these conditions.

We quantified the growth rates of all Hsp90 single-mutant variants in each of the stress conditions as selection coefficients (Figure 3.2—source data 1, Supplementary

Figures 3.6–3.10). We could clearly differentiate between the selection coefficients of wild-type synonyms and stop codons in all conditions (Figure 3.2B, Supplementary Figure 3.11) and we normalized to these classes of mutations to facilitate comparisons between each condition (Supplementary Figure 3.12). Of note, the observed selection coefficients of wild-type synonyms varied more in conditions of high temperature and diamide stress compared to standard (Supplementary Figures 3.13A,B). We also note greater variation in the selection coefficients of barcodes for the same codon in the diamide and high temperature conditions (Supplementary Figure 3.13C). We conclude that diamide and elevated temperature provided greater noise in our selection coefficient measurements. To take into account differences in signal to noise for each condition, we either averaged over large numbers of mutations or categorized selection coefficients as wild-type-like, strongly deleterious, intermediate, or beneficial based on the distribution of wild-type synonyms and stop codons in each condition (see Materials and methods and Supplementary Figure 3.12).

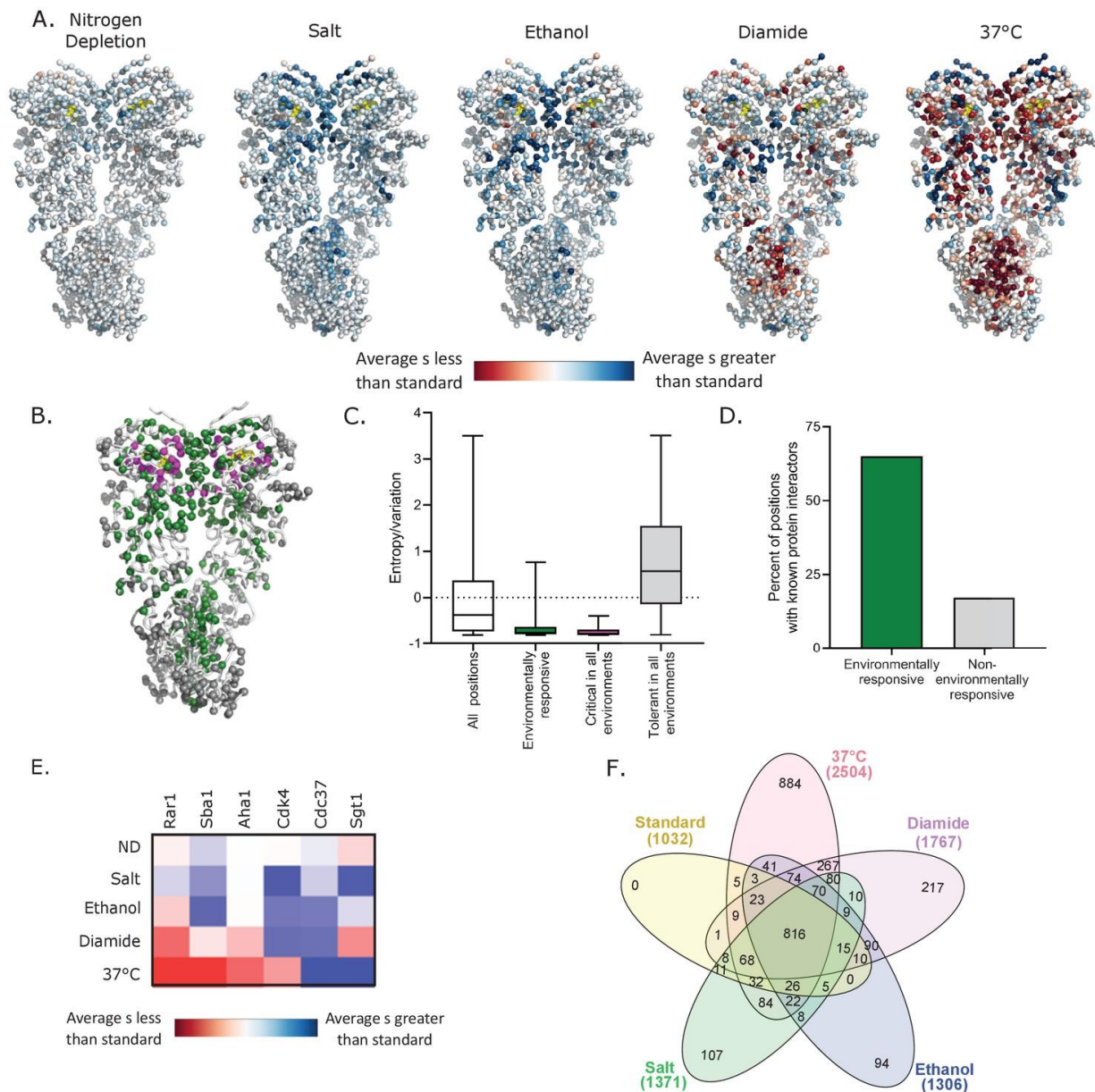
We compared selection coefficients of each Hsp90 variant in each stress condition to standard condition (Figure 3.2B&C). The stresses of 37°C and diamide tend to exaggerate the growth defects of many mutants compared to standard conditions, whereas, high salt and ethanol tend to rescue growth defects (Figure 3.2B&C and Supplementary Figure 3.14). According to the theory of metabolic flux (Kacser and Burns 1981; Dykhuizen et al. 1987), gene products that are rate limiting for growth will be subject to the strongest selection. Accordingly, the relationship between Hsp90 function and growth rate should largely determine the strength of selection acting on Hsp90 sequence. Conditions where Hsp90 function is more directly linked to growth

rate would be more sensitive to Hsp90 mutations than conditions where Hsp90 function can be reduced without changing growth rates (Bershtein et al. 2013; Jiang et al. 2013). The average selection coefficients are more deleterious in diamide and temperature stress compared to standard conditions. These findings are consistent with heat and diamide stresses causing a growth limiting increase in unfolded Hsp90 clients. In contrast, the average selection coefficients are less deleterious in ethanol and salt stress than in standard conditions, which suggests a decrease in the demand for Hsp90 function in these conditions. Due to the complex role Hsp90 plays in diverse signaling pathways in the cell, the different environmental stresses may differentially impact subsets of client proteins that cause distinct selection pressures on Hsp90 function.

### **Structural Analyses of Environmental Responsive Positions**

Altering environmental conditions had a pervasive influence on mutational effects along the sequence of Hsp90 (Figure 3.3A and Supplementary Figure 3.15). We structurally mapped the average selection coefficient of each position in each condition relative to standard conditions as a measure of the sensitivity to mutation of each position under each environmental stress (Figure 3.3A, Figure 3.3—source data 1). Many positions had mutational profiles that were responsive to a range of environments. Environmentally responsive positions with large changes in average selection coefficient in at least three conditions are highlighted on the Hsp90 structure in green in (Figure 3.3B). Unlike the critical positions that cluster around the ATP binding site (Figure 3.1C), the environmentally responsive positions are located throughout all domains of Hsp90. Similar to critical residues, environmentally responsive positions are more conserved in nature compared to other positions in Hsp90 (Figure 3.3C),

suggesting that the suite of experimental stress conditions tested captured aspects of natural selection pressures on Hsp90 sequence.



**Figure 3.3 Environmental stresses place distinct selection pressures on Hsp90.**

(A) The average selection coefficient ( $s$ ) at each position relative to standard conditions was mapped onto Hsp90 structure (Ali et al. 2006)(See Figure 3—source data 1). (B) Structural images indicating the location of positions that are critical for Hsp90 function

in all conditions (magenta), positions that are environmentally responsive (ER) (green), and positions that are tolerant in all environments (gray). Critical residues have mean selection coefficients that are null-like in all environments. ER positions have mean selection coefficients that differed from standard in three or more environments by an amount greater than one standard deviation of wild-type synonyms. Tolerant residues are not shifted more than this cutoff in any environment. **(C)** For different classes of positions, evolutionary variation was calculated as amino acid entropy at each position in Hsp90 sequences from diverse eukaryotes. Distributions are significantly different as measured by a two-sample Kolmogorov-Smirnov (KS) (All positions vs. ER: N = 678, 137,  $p < 0.0001$ , D = 0.39; All positions vs. critical: N = 678, 27,  $p < 0.0001$ , D = 0.57; All positions vs. tolerant: N = 678, 136,  $p < 0.0001$ , D = 0.38) **(D)** Fraction of different classes of mutations located at contact sites with binding partners.  $p < 0.0001$  **(E)** A heatmap of the average selection coefficient for all positions at the stated interfaces relative to standard conditions in each environment. **(F)** Venn diagram of deleterious mutations in different environmental conditions (Heberle et al. 2015). Total number of deleterious mutants in each condition are stated in parentheses.

### Figure 3—source data 1

**Average selection coefficient (excluding stops) at each position of Hsp90 in each environmental condition relative to the average selection coefficient in standard conditions.**

<https://cdn.elifesciences.org/articles/53810/elifesciences-53810-fig3-data1-v2.xlsx>

[Download elifesciences-53810-fig3-data1-v2.xlsx](#)

Hsp90 positions with environmentally responsive selection coefficients were enriched in binding contacts with clients, co-chaperones and intramolecular Hsp90 contacts involved in transient conformational changes (Figure 3.3D and Supplementary Figure 3.16A). About 65% of the environmentally responsive residues have been identified either structurally or genetically as interacting with binding partners (Bohen and Yamamoto 1993; Nathan and Lindquist 1995; Meyer et al. 2003; Meyer et al. 2004;

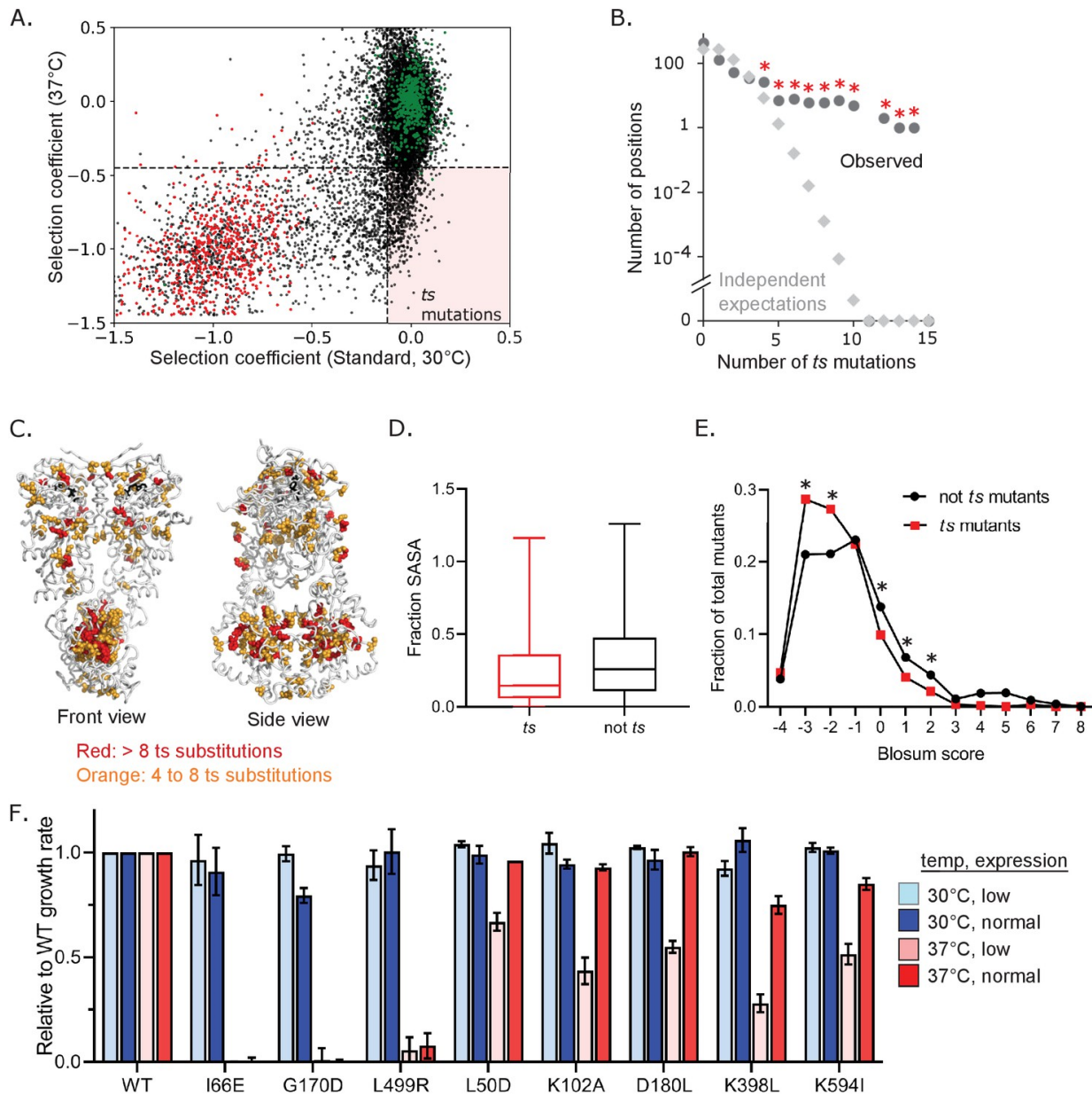


Roe et al. 2004; Ali et al. 2006; Hawle et al. 2006; Retzlaff et al. 2009; Zhang et al. 2010; Hagn et al. 2011; Genest et al. 2013; Lorenz et al. 2014; Verba et al. 2016; Kravats et al. 2018), compared to about 15% of positions that were not responsive to stress conditions. This analysis was performed on the small subset of clients and cochaperones with known Hsp90-binding sites. While ATP binding and hydrolysis are the main structural determinants that constrain fitness in standard growth conditions, client and co-chaperone interactions have a larger impact on experimental fitness under stress conditions. Although the mean selection coefficients of mutations at the known client and co-chaperone binding sites are responsive to changes in environment, the direction of the shift of growth rate compared to standard conditions depends on the specific binding partner and environment (Figure 3.3E and Supplementary Figure 3.16B). This suggests that different environments place unique functional demands on Hsp90 that may be mediated by the relative affinities of different clients and co-chaperones. Consistent with these observations, we hypothesize that Hsp90 client priority is determined by relative binding affinity and that Hsp90 mutations can reprioritize clients that in turn impacts many signaling pathways.

### **Constraint of Mutational Sensitivity at High Temperature**

We find that different environmental conditions lead to distinct selection on Hsp90 based on the number of beneficial and deleterious variants in each condition, including elevated temperature placing the greatest purifying selection pressure on Hsp90. Of the 2504 variants of Hsp90 that are deleterious when grown at 37°C, 884 of them (~35%) are deleterious only in this condition (Figure 3.3F). We defined mutants

that confer temperature sensitive (*ts*) growth phenotypes on cells as variants with selection coefficients within the distribution of wild-type synonyms in standard conditions and that of stop codons at 37°C. Based on this definition, 663 Hsp90 amino acid changes (roughly 5% of possible changes) were found to be temperature sensitive (Figure 3.4A, Figure 3.4—source data 1). We sought to understand the physical underpinnings of this large set of Hsp90 *ts* mutations.



**Figure 3.4: Abundance and mechanism of temperature-sensitive (*ts*) mutations in Hsp90.** (A) *Ts* variants were identified that supported WT-like growth at 30°C, but were null-like at 37°C in bulk competitions. WT synonyms are shown in green and stops in red. The horizontal dashed line corresponds to  $y = -0.47$ , the upper limit of stops at 37°C and the vertical dashed line corresponds to  $x = -0.11$ , the lower limit of WT synonyms in standard conditions. The pink-shaded quadrant highlights *ts* mutations. All *ts* mutants are listed in Figure 4—source data 1. (B) Distribution of the number of observed *ts* mutations at the same positions of Hsp90 (●) is much greater than

expected if they had occurred independently ( $\diamond$ ). Independent expectations were calculated as the probability of the stated number of mutations occurring at the same position by chance. \* Indicates observations that were significantly ( $p < 0.01$ ) greater than independent expectations based on random simulations and one-tailed t-tests. **(C)** Mapping positions with multiple *ts* variants onto Hsp90 structure. ATP is shown in black. **(D)** Solvent accessible surface area (SASA) of *ts* mutants compared to non-*ts* mutants. A two sample KS test showed significant differences in distributions ( $N = 663, 11762, p < 0.0001, D = 0.1735$ ) **(E)** Amino acid similarity to the wild type was estimated as the Blosum score (Henikoff and Henikoff 1992) for *ts* and non-*ts* variants. A two-proportion z-test was performed on each pair for each Blosum score and their p-values were adjusted using Benjamini-Hochberg adaptive step-up procedure. \*Indicates  $p < 0.05$ . **(F)** Growth rate of a panel of individual Hsp90 *ts* variants analyzed in isolation.

**Figure 3.4—source data 1: List of all temperature-sensitive mutants and associated selection coefficients.** *Ts* mutants were defined as variants with selection coefficients within the distribution of wild-type synonyms in standard conditions and that of stop codons at 37°C.

<https://cdn.elifesciences.org/articles/53810/elife-53810-fig4-data1-v2.xlsx>

[Download elife-53810-fig4-data1-v2.xlsx](#)

We examined Hsp90 *ts* mutations for structural and physical patterns. We found that *ts* mutations tended to concentrate at certain amino acid positions of Hsp90 (Figure 3.4B). The clustering of *ts* mutations was significant compared to random simulations. Positions with greater than four *ts* mutations were spread across all three domains of Hsp90 (Figure 3.4C) with the largest cluster occurring in the C domain of Hsp90. The C domain forms a constitutive homodimer that is critical for function (Wayne and Bolon 2007). Of note, homo-oligomerization domains may have a larger *ts* potential because all subunits contribute to folding and dimerization essentially multiplying the impacts of mutations (Lynch 2013). To explore the physical underpinnings of *ts* mutations, we

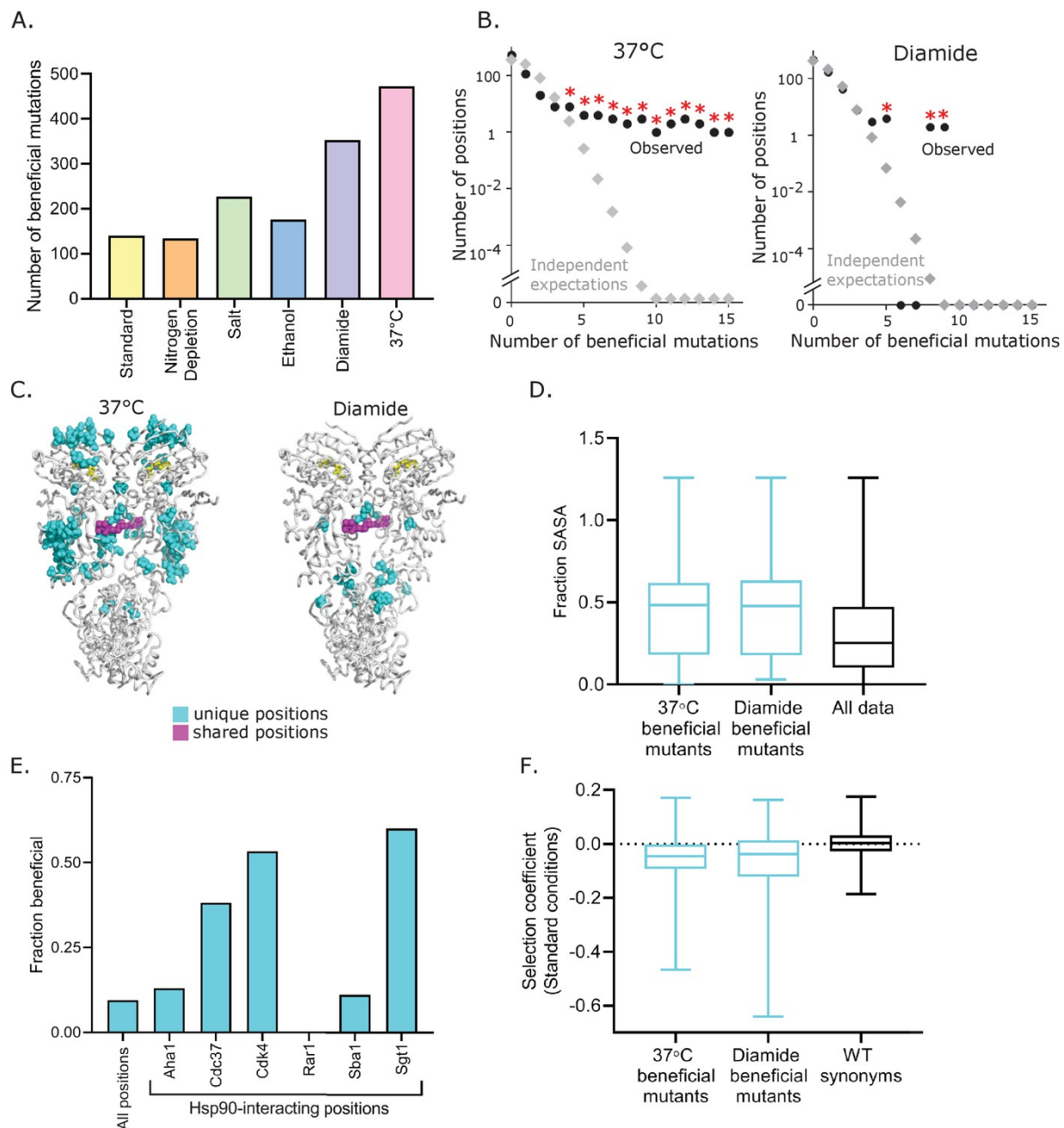
examined if they were buried in the structure or surface exposed. Mutations at buried residues tend to have a larger impact on protein folding energy compared to surface residues (Chakravarty S 1999). Consistent with the idea that many *ts* mutations may disrupt protein folding at elevated temperature, substitutions that confer a *ts* phenotype are enriched in buried residues (Figure 3.4D). Also consistent with this idea, *ts* mutations tend to have negative Blosum scores (Henikoff and Henikoff 1992); (Figure 3.4E), a hallmark of disruptive amino acid changes.

Because growth at elevated temperatures requires higher levels of Hsp90 protein (Borkovich et al. 1989), some *ts* mutations are likely due to a reduced function that is enough for growth at standard temperature, but is insufficient at 37°C (Nathan and Lindquist 1995). We reasoned that we could distinguish these mutants by examining how growth rate depended on the expression levels of Hsp90. We expect that destabilizing mutants that cause Hsp90 to unfold at elevated temperature would not support efficient growth at 37°C independent of expression levels. In contrast, we expect mutants that reduce Hsp90 function to exhibit an expression-dependent growth defect at 37°C. We tested a panel of *ts* mutations identified in the bulk competitions at high and low expression levels (Figure 3.4F). The dependence of growth rate at 37°C on expression level varied for different Hsp90 *ts* variants. The I66E, G170D and L499R Hsp90 mutants have no activity at 37°C irrespective of expression levels. These disruptive substitutions at buried positions likely destabilize the structure of Hsp90. In contrast, increasing the Hsp90 expression levels at least partially rescued the growth defect for five *ts* variants (L50D, K102A, D180L, K398L, K594I), indicating that these variants do not provide enough Hsp90 function for robust growth at elevated

temperature. All five of these expression-dependent *ts* variants were located at surface positions. Thus, for the *ts* mutants we tested individually, we see a correlation between location of the mutation and type of *ts* mutation. Destabilizing mutations tend to be buried and mutants with reduced function tend to be surface exposed, indicating that the location of *ts* mutations can delineate these different mechanistic classes.

### **Hsp90 Potential for Adaptation to Environmental Stress**

Numerous Hsp90 variants provided a growth benefit compared to the wild-type sequence in stress conditions. The largest number of beneficial variants in Hsp90 occurred in high temperature and diamide conditions (Figure 3.5A, Figure 3.5—source data 1). Multiple lines of evidence indicate that these mutants are truly beneficial variants and not simply measurement noise. First, the beneficial amino acids generally exhibited consistent selection coefficients among synonymous variants (Supplementary Figure 3.17A). Second, beneficial mutants in diamide and high temperature tend to cluster at certain positions (Figure 3.5B), which would not be expected for noise. Finally, we confirmed the increased growth rate at elevated temperature of a panel of variants analyzed in isolation (Supplementary Figure 3.17B). The fact that beneficial mutations in elevated temperature and diamide often clustered at specific positions in Hsp90 indicates that the wild-type amino acids at these positions are far from optimum for growth in these conditions. In contrast, the apparent beneficial mutations in other conditions did not tend to cluster at specific positions (Supplementary Figure 3.18).



**Figure 3.5: Beneficial variants in diamide and elevated temperature conditions.**

(A) Number of beneficial mutations identified in each condition based on selection coefficients more than two standard deviations greater than wild-type synonyms. Beneficial mutants at 37°C and in diamide are listed in (Figure 3.5—source data 1). (B) Distribution of the number of beneficial mutations at the same position in both 37°C (left) and diamide (right) conditions (●) is greater than expected if they had occurred independently (◇). Independent expectations were calculated as the probability of the stated number of mutations occurring at the same position by chance. \* Indicates

observations that were significantly ( $p < 0.01$ ) greater than independent expectations based on random simulations and one-tailed t-tests. **(C)** Location of positions with four or more beneficial mutations. Positions that are unique to diamide or 37°C are shown in cyan and two shared positions are shown in magenta. **(D)** The solvent accessible surface area (SASA) of beneficial mutations at 37°C and in diamide compared to all mutations. Distributions are significantly different as measured by a two-sample KS test (37°C vs. all data,  $N = 270, 12393$ ,  $p < 0.0001$ ,  $D = 0.2851$ ; Diamide vs. all data,  $N = 60, 12393$ ,  $p < 0.0001$ ,  $D = 0.3465$ ) **(E)** The fraction of Hsp90 positions at interfaces that were beneficial in 37°C and diamide conditions. **(F)** Selection coefficients in standard conditions for beneficial mutations at 37°C and in diamide compared to wild-type synonyms. (KS test; 37°C vs. WT synonyms:  $N = 463, 660$ ,  $p < 0.0001$ ,  $D = 0.3281$ ; Diamide vs. WT synonyms:  $N = 353, 660$ ,  $p < 0.0001$ ,  $D = 0.3809$ ).

**Figure 3.5—source data 1: List of all beneficial mutants and associated selection coefficients at 37°C and in diamide.** Beneficial mutants were defined as variants with selection coefficients two standard deviations greater than wild-type synonyms. <https://cdn.elifesciences.org/articles/53810/elifesciences-53810-fig5-data1-v2.xlsx> [Download elifesciences-53810-fig5-data1-v2.xlsx](#)

To obtain a more general picture of the potential for adaptation derived from the full fitness distributions, we used Fisher's Geometric model (FGM) (Fisher 1931).

According to FGM, populations evolve in an  $n$ -dimensional phenotypic space, through random single-step mutations, and any such mutation that brings the population closer to the optimum is considered beneficial. An intuitive hypothesis derived from FGM is that the potential for adaptation in a given environment (i.e. is the availability of beneficial mutations) depends on the distance to the optimum. In order to estimate the distance to the optimum  $d$ , we adopted the approach by Martin and Lenormand and fitted a displaced gamma distribution to the neutral and beneficial mutations for each environment (Martin and Lenormand 2006). We observed that the yeast populations were furthest from the optimum in elevated temperature and diamide ( $d = 0.072$  and  $0.05$ , respectively), followed by nitrogen deprivation ( $d = 0.023$ ), high salinity and



ethanol ( $d = 0.021$ ) and standard ( $d = 0.014$ ). This suggests that exposure to elevated temperature and diamide results in the largest potential for adaptation and is consistent with the observation of the largest proportions of beneficial mutations in these environments. Interestingly, previous results from a 9-amino-acid region in Hsp90 indicated that there was very little potential for adaptation at high temperature (36°C) as compared with high salinity (Hietpas et al. 2013). This apparent contradiction between results from the full Hsp90 sequence and the 582–590 region indicates that a specific region of the protein may be already close to its functional optimum in a specific environment, whereas, there is ample opportunity for adaptation when the whole protein sequence is considered.

In diamide and elevated temperature, the clustered beneficial positions were almost entirely located in the ATP-binding domain and the middle domain (Figure 3.5C), both of which make extensive contacts with clients and co-chaperones (Meyer et al. 2003; Meyer et al. 2004; Roe et al. 2004; Ali et al. 2006; Zhang et al. 2010; Verba et al. 2016). Beneficial mutations in elevated temperature and diamide conditions were preferentially located on the surface of Hsp90 (Figure 3.5D) at positions accessible to binding partners. Analyses of available Hsp90 complexes indicate that beneficial positions were disproportionately located at known interfaces with co-chaperones and clients (Figure 3.5E). Clustered beneficial mutations are consistent with disruptive mechanisms because a number of different amino acid changes can lead to disruptions, whereas, a gain of function is usually mediated by specific amino acid changes. Amino acids that are beneficial in diamide and elevated temperature tend to exhibit deleterious effects in standard conditions (Figure 3.5F), consistent with a cost of adaptation. In

comparison, wild-type-like mutations in diamide and high temperature tend to exhibit wild-type-like fitness in standard conditions (Supplementary Figure 3.19). We conjecture that the clustered beneficial mutations are at positions that mediate the binding affinity of subsets of clients and co-chaperones and that disruptive mutations at these positions can lead to re-prioritization of multiple clients. The priority or efficiency of Hsp90 for sets of clients can in turn impact most aspects of physiology because Hsp90 clients include hundreds of kinases that influence virtually every aspect of cell biology.

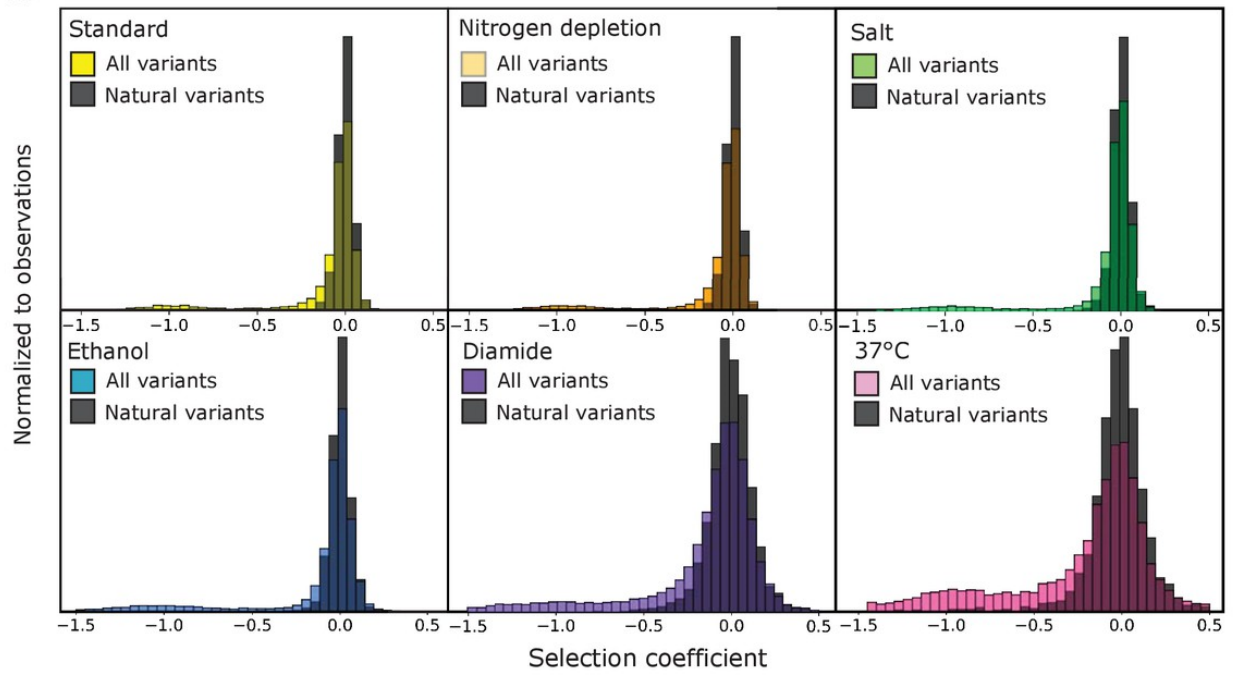
In the first seven amino acids of Hsp90, we noted both a large variation in the selection coefficients of synonymous mutations at elevated temperature and that many nonsynonymous substitutions at these positions generated strong beneficial effects (Supplementary Figure 3.20 A,B). Synonymous mutations at these positions were only strongly beneficial at high temperature where Hsp90 protein levels are limiting for growth. Analysis of an individual clone confirms that a synonymous mutation at the beginning of Hsp90 that was beneficial at high temperature was expressed at higher level in our plasmid system (Supplementary Figure 3.20C,D). These results are consistent with a large body of research showing that mRNA structure near the beginning of coding regions often impacts translation efficiency (Tuller et al. 2010; Plotkin and Kudla 2011; Li 2015), and that adaptations can be mediated by changes in expression levels (Lang et al. 2013).

### **Natural Selection Favors Hsp90 Variants That are Robust to Environment**

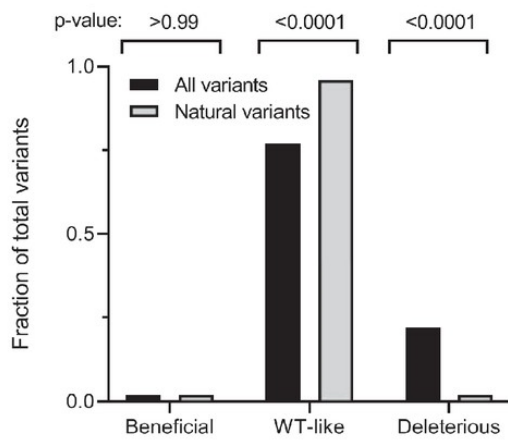
We next examined how experimental protein fitness maps compared with the diversity of Hsp90 sequences in current eukaryotes. We analyzed Hsp90 diversity in a

set of 267 sequences from organisms that broadly span across eukaryotes. We identified 1750 amino acid differences in total that were located at 499 positions in Hsp90. We examined the experimental growth effects of the subset of amino acids that were observed in nature. While the overall distribution of selection coefficients in all conditions was bimodal with peaks around neutral ( $s = 0$ ) and null ( $s = -1$ ), the natural amino acids were unimodal with a peak centered near neutral (Figure 3.6A, Figure 3.6—source data 1). The vast majority of natural amino acids had wild-type-like fitness in all conditions studied here (Figure 3.6B and C). Whereas, naturally occurring amino acids in Hsp90 were rarely deleterious in any experimental condition, they were similarly likely to provide a growth benefit compared to all possible amino acids (5%). This observation indicates that condition-dependent fitness benefits are not a major determinant of natural variation in Hsp90 sequences. Instead, our results indicate that natural selection has favored Hsp90 substitutions that are robust to multiple stressful conditions (Figure 3.6D). Beneficial mutations in heat and diamide indicate that there is room for improvement in Hsp90 function in individual conditions. The clustering of beneficial mutations at known binding interfaces suggests that the optimal binding affinity for partner proteins may depend on growth conditions. We propose that natural variants of Hsp90 have been selected for binding properties that are robust to different stresses rather than specific to individual conditions.

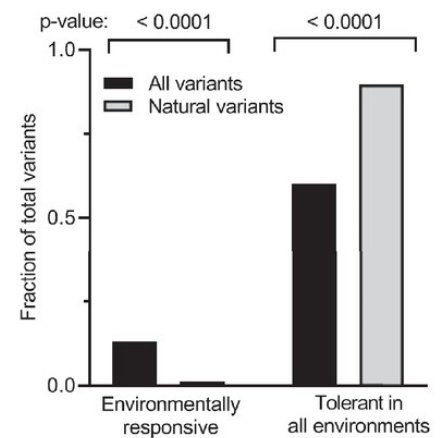
A.



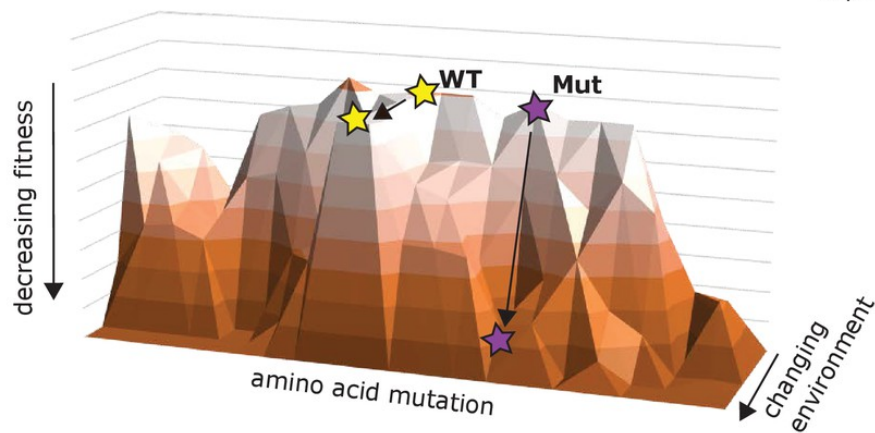
B.



C.



D.



**Figure 3.6: Experimental growth effects of natural amino acid variants of Hsp90.**

(A) The distribution of selection coefficients of natural variants compared to all variants in each environmental condition (see Figure 6—source data 1). (B) Across all environments, the fraction of natural variants compared to all variants that were beneficial, wild-type-like, or deleterious. (C) The fraction of natural variants compared to all variants that were environmentally responsive or tolerant in all environments. Categories were defined as in (Figure 3B). (D) Landscape model indicating that natural variants of Hsp90 tend to support robust growth under a variety of stress conditions.

**Figure 3.6—source data 1**

**List of selection coefficients of all-natural variants of Hsp90 in all environmental conditions.**

<https://cdn.elifesciences.org/articles/53810/elifesciences-53810-fig6-data1-v2.xlsx>

[Download elifesciences-53810-fig6-data1-v2.xlsx](#)

Epistasis may provide a compelling explanation for the naturally occurring amino acids that we observed with deleterious selection coefficients. Analyses of Hsp90 mutations in the context of likely ancestral states has demonstrated a few instances of historical substitutions with fitness effects that depend strongly on the Hsp90 sequence background (Starr et al., 2018). Indeed, many of the natural amino acids previously identified with strong epistasis (E7A, V23F, T13N) are in the small set of natural amino acids with deleterious effects in at least one condition. Further analyses of natural variants under diverse environmental conditions will likely provide insights into historical epistasis and will be the focus of future research.

**Discussion**

In this study, we analyzed the protein-wide distribution of fitness effects of Hsp90 across standard and five stress conditions. We found that environment has a profound effect on the fates of Hsp90 mutations. Each environmental stress varies in the strength of selection on Hsp90 mutations; heat and diamide increase the strength of selection

and ethanol and salt decrease the strength of selection. While proximity to ATP is the dominant functional constraint in standard conditions, the influence of client and co-chaperone interactions on growth rate dramatically increases under stress conditions. Additionally, beneficial mutations cluster at positions that mediate binding to clients and cochaperones. The fact that different Hsp90 binding partners have distinct environmental dependencies suggests that Hsp90 can reprioritize clients that in turn impacts many downstream signaling pathways.

Our results demonstrate that mutations to Hsp90 can have environment-dependent effects that are similar to the stress-induced changes to the function of wild-type Hsp90 that have been shown to contribute to new phenotypes (Jarosz and Lindquist 2010). The low frequency of environment-dependent amino acids in Hsp90 from extant eukaryotes indicates that this type of evolutionary mechanism is rare relative to drift and other mechanisms shaping Hsp90 sequence diversity.

We observed distinct structural trends for mutations that provide environment-dependent costs and benefits. Many mutations in Hsp90 caused growth defects at elevated temperature where Hsp90 function is limiting for growth. These temperature-sensitive mutations tended to be buried and in the homodimerization domain, consistent with an increased requirement for folding stability at elevated temperatures. In contrast, beneficial mutations tended to be on the surface of Hsp90 and at contact sites with binding partners, suggesting that change-of-function mutations may be predominantly governed by alterations to binding interactions. Mutations that disrupt binding to certain clients can lead to the re-prioritization of others, which, due to Hsp90's central role in numerous cellular pathways, has the potential to modify integral networks in response

to stress. Once more comprehensive data is available on Hsp90-client binding sites, it may be possible to simulate this rewiring of cellular networks, providing insight into the causes of the beneficial mutations. However, presently, the large number of clients with unknown binding sites makes these analyses challenging. In the future, comparing Hsp90 client-interactomes (Taipale et al. 2012) may help delineate adaptive biochemical mechanisms.

### **Limitations:**

Our experimental setup has limitations that we have tried to account for in our analyses and conclusions. For example, we measured the fitness effects of Hsp90 under artificially low expression where yeast growth rates are tightly coupled to function of Hsp90 in order to provide a sensitive readout of fitness defects (Jiang et al. 2013). Expression of Hsp90 under this promoter remains stable in the stresses tested (Supplementary Figure 3.21). However, this defined promoter does not capture the native transcriptional regulation and may not fully recapitulate translational and post-translational regulation controlling *hsp82*. While these levels of regulation of Hsp90 are clearly important physiologically, the sensitive readouts of fitness that we measured appear to capture critical features of Hsp90 with regard to biochemical function and evolutionary mechanism. For example, virtually all deleterious mutants measured in this study under stress conditions appear to have also been subject to purifying selection in nature.

In addition, the experimental strain used in this study is deleted for the constitutively expressed paralog of *hsp82*, *hsc82*. *Hsp82* and *hsc82* are functionally

overlapping, essential genes with 97% sequence identity (16 amino acid differences) that can compensate for each other's loss-of-function in normal growth conditions (Girstmair et al. 2019). The high sequence identity between the two paralogs indicates that they are both under similar selection pressure. Despite the high sequence identity, a number of distinct differences have been noted in stability, conformational cycles, and client interactomes. Experimental evidence indicates that *hsp82* is more stress-specific, and more stable to unfolding (Girstmair et al. 2019). Further efforts will be required to resolve how distinctions between Hsp90 paralogs contribute to function and selection.

### **Relationship to Prior Work:**

A handful of studies have assessed the impact of environment on the fitness landscape of genes (Hietpas et al. 2013; Mavor et al. 2016; Dandage et al. 2017; Li and Zhang 2018). For example, Dandage, Chakraborty and colleagues investigated the effects of temperature and chemical chaperones on the fitness landscape of the Gentamicin resistance gene and found that protein stability and distance to the ligand binding site are the molecular properties with the strongest correlations with fitness (Dandage et al. 2017). To understand the strength of the molecular constraints on Hsp90 on a whole protein level, we performed similar analyses (Supplementary Figure 3.22). Consistent with the Gentamicin study, we find the features that best correlate with fitness are protein stability and distance from the active site. The constraint of protein stability is the highest at 37°C, indicating increased dependence on stability at the higher temperature. In addition, distance from the ATP-binding site imposes strong molecular constraints on Hsp90, signifying the importance of ATP hydrolysis on Hsp90 function. While individual



features correlate with fitness effects and show environmental dependence, single features are unable to capture the majority of observed variance in fitness effects, consistent with a complex set of physical properties that underlie fitness effects in both proteins.

In another study of the effect of environment on mutational fitness, Li and Zhang detected pervasive genotype-by-environment interactions between a yeast tRNA gene and environment (Li and Zhang 2018). They found that the correlation of the fitness between mutations in each tested environment was linear such that the fitness landscape in one environment together with a change in slope could be used to accurately predict fitness effects in the second environment. In this study, we observed a large impact of environment on Hsp90 fitness; however, we observe many fitness effects that deviate from a linear relationship between environments. While linear models can predict the fitness of some mutations in different environments, it would not predict many of the types of mutations that are focuses of this study, such as mutations that exhibit an adaptive trade-off, those with beneficial effects in one environment that become deleterious in another. In addition, the linear model would not predict the large group of *ts* mutations with wild-type fitness in standard conditions and null fitness at 37°C. As environmental-dependent protein fitness landscapes are analyzed for an increasing set of genes, it will provide opportunities to explore how different protein properties such as the number of binding partners may contribute to global trends.

Importantly, our results demonstrate that while mutations to Hsp90 can provide a growth advantage in specific environmental conditions, naturally occurring amino acids

in Hsp90 tend to support robust growth over multiple stress conditions. The finding of beneficial mutations in Hsp90 in specific conditions suggests that similar long-term stresses in nature can lead to positive selection on Hsp90. Consistent with previous work (Hietpas et al. 2013), we found that experimentally beneficial mutations tended to have a fitness cost in alternate conditions (Figure 3.5F). This indicates that natural environments which fluctuate among different stresses would reduce or eliminate positive selection on Hsp90. Therefore, our results suggest that natural selection on Hsp90 sequence has predominantly been governed by strong purifying selection integrated over multiple stressful conditions. Taken together, these results support the hypothesis that natural populations might experience a so-called ‘micro-evolutionary fitness seascape’ (Mustonen and Lässig 2009), in which rapidly fluctuating environments result in a distribution of quasi-neutral substitutions over evolutionary time scales.

## **Materials and Methods**

### **Generating Mutant Libraries**

A library of Hsp90 genes was saturated with single point mutations using oligos containing NNN codons as previously described (Hietpas et al. 2012). The resulting library was pooled into 12 separate 60 amino acid long sub-libraries (amino acids 1–60, 61–120 etc.) and combined via Gibson Assembly (NEB) with a linearized p414ADH $\Delta$ ter Hsp90 destination vector, a low copy number plasmid with the *trp1* selectable marker. To simplify sequencing steps during bulk competition, each variant of the library was tagged with a unique barcode. For each 60 amino acid sub-library, a pool of DNA

constructs containing a randomized 18 bp barcode sequence (N18) was cloned 200 nt downstream from the Hsp90 stop codon via restriction digestion, ligation, and transformation into chemically competent *E. coli* with the goal of each mutant being represented by 10–20 unique barcodes.

### **Barcode Association of Library Variants**

We added barcodes and associated them with Hsp90 variants essentially as previously described (Starr et al. 2018). To associate barcodes with Hsp90 variants, we performed paired-end sequencing of each 60 amino acid sub-library using a primer that reads the N18 barcode in one read and a primer unique to each sub-library that anneals upstream of the region containing mutations. To facilitate efficient Illumina sequencing, we generated PCR products that were less than 1 kb in length for sequencing. We created shorter PCR products by generating plasmids with regions removed between the randomized regions and the barcode. To remove regions from the plasmids, we performed restriction digest with two unique enzymes, followed by blunt ending with T4 DNA polymerase (NEB) and plasmid ligation at a low concentration (3 ng/ $\mu$ L) to favor circularization over bimolecular ligations. The resulting DNA was re-linearized by restriction digest and amplified with 11 cycles of PCR to generate products for Illumina sequencing. The resulting PCR products were sequenced using an Illumina MiSeq instrument with asymmetric reads of 50 bases for Read1 (barcode) and 250 bases for Read2 (Hsp90 sequence). After filtering low-quality reads (Phred scores < 10), the data was organized by barcode sequence. For each barcode that was read more than three

times, we generated a consensus of the Hsp90 sequence that we compared to wild type to call mutations. Of note, the building of consensus of at least three independent reads reduces the chance that errors will lead to mistaken variant identity because the same misread would have to occur in the majority of these reads.

### **Bulk Growth Competitions**

Equal molar quantities of each sub-library were mixed to form a pool of DNA containing the entire Hsp90 library with each codon variant present at similar concentration. The plasmid library was transformed using the lithium acetate procedure into the DBY288 Hsp90 shutoff strain of *S. cerevisiae* which has both genomic paralogs of Hsp90 (*hsp82* and *hsc82*) deleted and a chromosomal copy of *hsp82* under a galactose-dependent promoter inserted (*can1-100 ade2-1 his3-11,15 leu2-3,12 trp1-1, ura3-1 hsp82::leu2 hsc82::leu2 ho::pgals-hsp82-his3*) essentially as previously described (Jiang et al. 2013). Sufficient transformation reactions were performed to attain ~5 million independent yeast transformants representing a fivefold sampling for the average barcode and 50 to 100-fold sampling for the average codon variant. Following 12 hr of recovery in SRGal (synthetic 1% raffinose and 1% galactose) media, transformed cells were washed five times in SRGal-W media (SRGal lacking tryptophan to select for the presence of the Hsp90 variant plasmid) to remove extracellular DNA, and grown in SRGal-W media at 30°C for 48 hr with repeated dilution to maintain the cells in log phase of growth. This yeast library was supplemented with 20% glycerol, aliquoted and slowly frozen in a -80°C freezer.

For each competition experiment, an aliquot of the frozen yeast library cells was thawed at 37°C. Viability of the cells was accessed before and after freezing and was determined to be greater than 90% with this slow freeze, quick thaw procedure. Thawed cells were amplified in SRGal-W for 24 hr, and then shifted to shutoff conditions by centrifugation, washing, and resuspension in 300 mL of synthetic dextrose lacking tryptophan (SD-W) for 12 hr at 30°C. At this time, cells containing a null-rescue plasmid had stopped growing and Hsp90 was undetectable by western blot (Supplementary Figure 3.23). At this point, cells were split and transferred to different conditions including: Standard (SD-W, 30°C), Nitrogen depletion (SD-W with limiting amounts of ammonium sulfate, 0.0125%, 30°C), Salt (SD-W with 0.8 M NaCl, 30°C), Ethanol (SD-W with 7.5% ethanol, 30°C), Diamide (SD-W with 0.85 mM diamide, 30°C), or high temperature (SD-W, 37°C). We collected samples of  $\sim 10^8$  cells at eight time points over a period of 36 hr and stored them at  $-80^\circ\text{C}$ . Cultures were maintained in log phase by regular dilution with fresh media every 6–10 hr to maintain a population size of  $10^8$ – $10^9$  cells in order to prevent population bottlenecks relative to sample diversity. Bulk competition from the standard condition were conducted in technical duplicates from the frozen yeast library.

### **DNA Preparation and Sequencing**

We isolated plasmid DNA from each bulk competition time point as described (Jiang et al. 2013). Purified plasmid was linearized with *Ascl*. Barcodes were amplified by 19 cycles of PCR using Phusion polymerase (NEB) and primers that add Illumina adapter sequences and an 8 bp identifier sequence used to distinguish libraries and time points. The identifier sequence was located at positions 91–98 relative to the

Illumina primer and the barcode was located at positions 1–18. PCR products were purified two times over silica columns (Zymo Research) and quantified using the KAPA SYBR FAST qPCR Master Mix (Kapa Biosystems) on a Bio-Rad CFX machine. Samples were pooled and sequenced on an Illumina NextSeq instrument in single-end 100 bp mode.

### **Analysis of Bulk Competition Sequencing Data**

Illumina sequence reads were filtered for Phred scores  $> 20$  and strict matching of the sequence to the expected template and identifier sequence. Reads that passed these filters were parsed based on the identifier sequence. For each condition/time-point identifier, each unique N18 read was counted. The unique N18 count file was then used to identify the frequency of each mutant using the variant-barcode association table. This barcoding strategy reduces the impact of bases misread by Illumina, as they result in barcodes that are not in our lookup table created by paired end sequencing and thus are discarded from the fitness analyses. To generate a cumulative count for each codon and amino acid variant in the library, the counts of each associated barcode were summed. To reduce experimental noise, selection coefficients were not calculated for variants with less than 100 reads at the 0 time point (Boucher et al. 2014). The average variant at the 0 time point had approximately 500 reads.

### **Determination of Selection Coefficient**

Selection coefficients were estimated using empiricIST (Fragata et al. 2018), a software package developed based on a previously published Markov Chain Monte Carlo (MCMC) approach (Bank et al. 2014). Briefly, we estimated individual growth

rates and initial population sizes relative to the wild-type sequence simultaneously, based on a model of exponential growth and multinomial sampling of sequencing reads independently at each time point. For each mutant we obtained 10,000 posterior samples for the growth rate and initial population using a Metropolis-Hastings algorithm. The resulting growth rate estimates correspond to the median of 1000 samples of the posterior. Subsequently, selection coefficients ( $s$ ) were scaled so that the average stop codon in each environmental condition represented a null allele ( $s = -1$ ). For the second replicate in standard conditions, we noted a small fitness defect ( $s \approx -0.2$ ) for wild-type synonyms at positions 679–709 relative to other positions. We do not understand the source of this behavior and chose to normalize to wild-type synonyms from 1 to 678 for this condition and to exclude positions 679–709 from analyses that include the second replicate of standard conditions. We did not observe this behavior in any other condition including the first standard condition replicate. Variants were categorized as having wild-type-like, beneficial, intermediate, or deleterious fitness based on the comparison of their selection coefficients with the distribution of wild-type synonyms and stop codons in each condition (Supplementary Figure 3.12) in the following manner; Wild-type-like: variants with selection coefficients within two standard deviations (SD) of the mean of wild-type synonyms; Beneficial: variants with selection coefficients above two SD of wild-type synonyms; Strongly deleterious: variants with selection coefficients within two SD of stop codons; Intermediate: variants with selection coefficients between those of stop-like and wild-type-like. Where stated, the average selection coefficient was calculated as the mean selection coefficient of all mutations at a position excluding that of the stop codon.

## Structural Analysis

The solvent accessible surface area was computed by the algorithm of (Lee and Richards 1971) using the PDB 2cg9 structure with the chains for Sba1 removed. The Blossom score was derived from the Blossom62 matrix (Henikoff and Henikoff 1992). Evolutionary conservation was calculated with an alignment of homologs from diverse species using the ConSurf server (Ashkenazy et al. 2016). The change in protein stability upon mutation ( $\Delta\Delta G$ ) was predicted by the PoPMuSic server (Dehouck et al. 2011). Distance from the  $\gamma$ -phosphate of ATP to the C- $\alpha$  of each amino acid residue was calculated using Pymol. Physico-chemical properties of the amino acids were retrieved from (Abriata et al. 2015). Correlation coefficients were calculated by Pearson product-moment correlations unless otherwise stated.

## Random Simulations to Assess Clustering of Mutations

To assess if classes of mutations (e.g. temperature-sensitive mutations) clustered at positions more than expected based on chance, we compared the observed distribution of mutations to random simulations. For the random simulations, we randomly selected a position for the number of observed mutations and stored the clustering distribution (e.g. the number of positions with 0, 1, 2, 3, etc. simulated mutations). We performed 1000 simulations and used the average and standard deviation from these simulations to define statistical cutoffs for random expectations.

## Yeast Growth Analysis

Individual variants of Hsp90 were generated by site-directed mutagenesis and confirmed by Sanger sequencing. Variants were cloned in a p414 plasmid either under



a low-expression, ADH promoter, or a high-expression, GPD promoter, as specified. Variants were generated by site directed mutagenesis and transformed into DBY288 cells. Selected transformed colonies were grown in liquid SRGal-W media to mid-log phase at 30°C, washed three times and grown in shutoff media (SD-W) for 10 hrs at 30°C, and then either kept at 30°C or shifted to 37°C as indicated. After sufficient time to stall the growth of control cells lacking a rescue copy of Hsp90 (~16 hr), cell density was monitored based on absorbance at 600 nm over time and fit to an exponential growth curve to quantify growth rate. Growth estimates were based on individual growth curves with at least four timepoints over an eight-hour period. Using this approach, we routinely observe measurement noise of 2–5%.

### **Analysis of Hsp90 Expression by Western Blot**

To analyze expression levels of Hsp90, cells were grown for the specified time in SD-W or the indicated environmental condition.  $10^8$  yeast cells were collected by centrifugation and frozen as pellets at  $-80^{\circ}\text{C}$ . Cells were lysed by vortexing the thawed pellets with glass beads in lysis buffer (50 mM Tris-HCl pH 7.5, 5 mM EDTA and 10 mM PMSF), followed by addition of 2% Sodium dodecyl sulfate (SDS). Lysed cells were centrifuged at 18,000 g for 1 min to remove debris, and the protein concentration of the supernatants was determined using a BCA protein assay kit (Pierce) compared to a Bovine Serum Albumin (BSA) protein standard. 15  $\mu\text{g}$  of total cellular protein was resolved by SDS-PAGE, transferred to a PVDF membrane, and Hsp90 was probed using an anti-human Hsp90  $\alpha/\beta$  antibody that cross reacts with yeast Hsp90 (Cayman chemical).

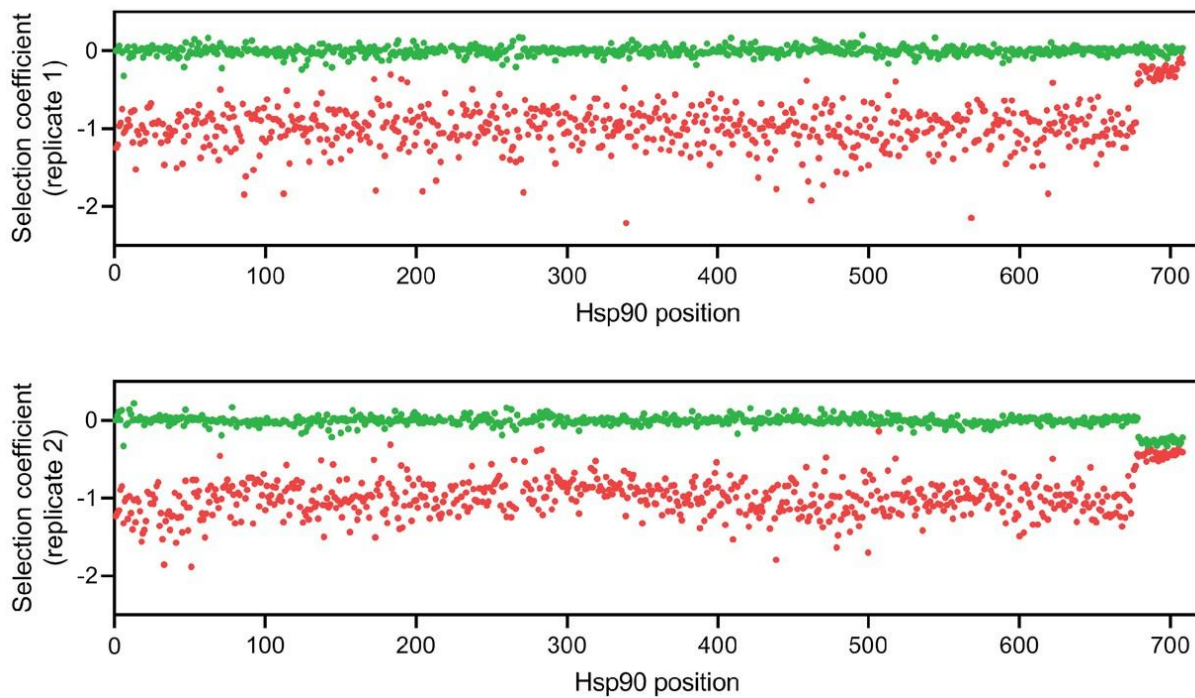
### **Natural Variation in Hsp90 Sequence**

We analyzed sequence variation in a previously described alignment of Hsp90 protein sequences from 261 eukaryotic species that broadly span a billion years of evolutionary distance (Starr et al. 2018)

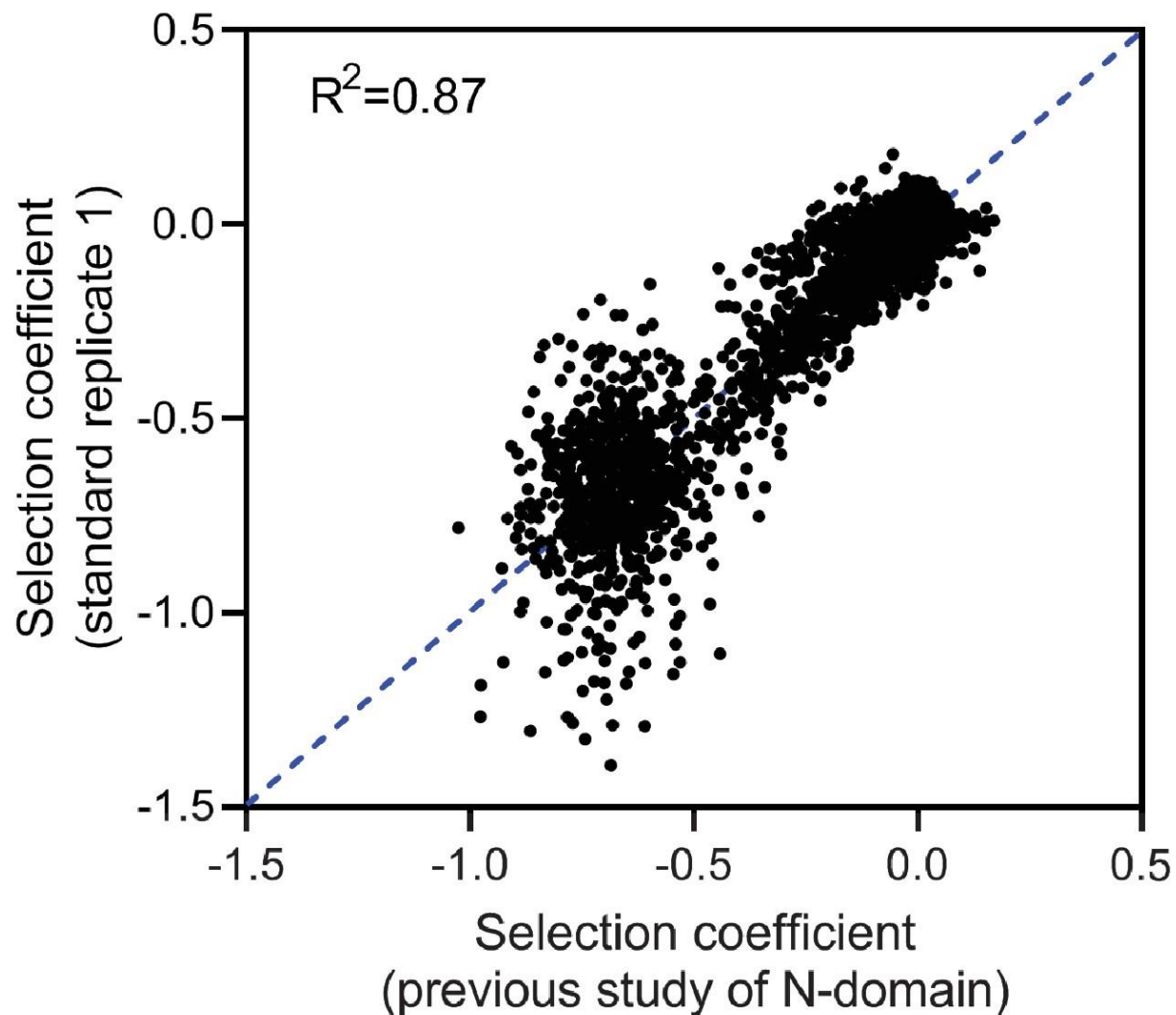
### **Acknowledgements**

Thanks to Tyler Starr for providing the alignment of Hsp90 sequences used to assess natural variation. This work was supported by grants from the National Institutes of Health (R01-GM112844 to DNAB and F32-GM119205 to JMF). IF was supported by a postdoctoral fellowship from the FCT (Fundação para a Ciência e a Tecnologia) within the project JPIAMR/0001/2016. CB is grateful for support from EMBO Installation Grant IG4152 and ERC Starting Grant 804569 - FIT2GO.

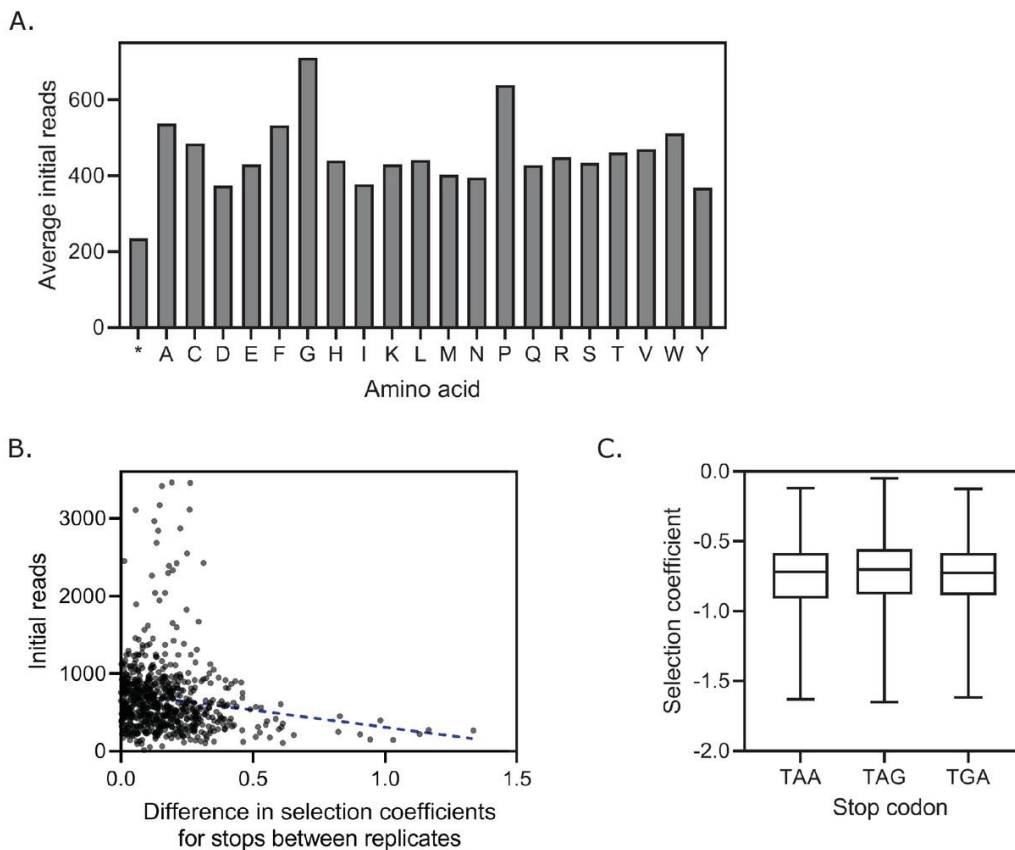
## Supplementary Figures



**Supplementary Figure 3.1: Selection coefficients for wild-type synonyms (green) and stops (red) at each position of Hsp90 for both replicates of standard conditions.**



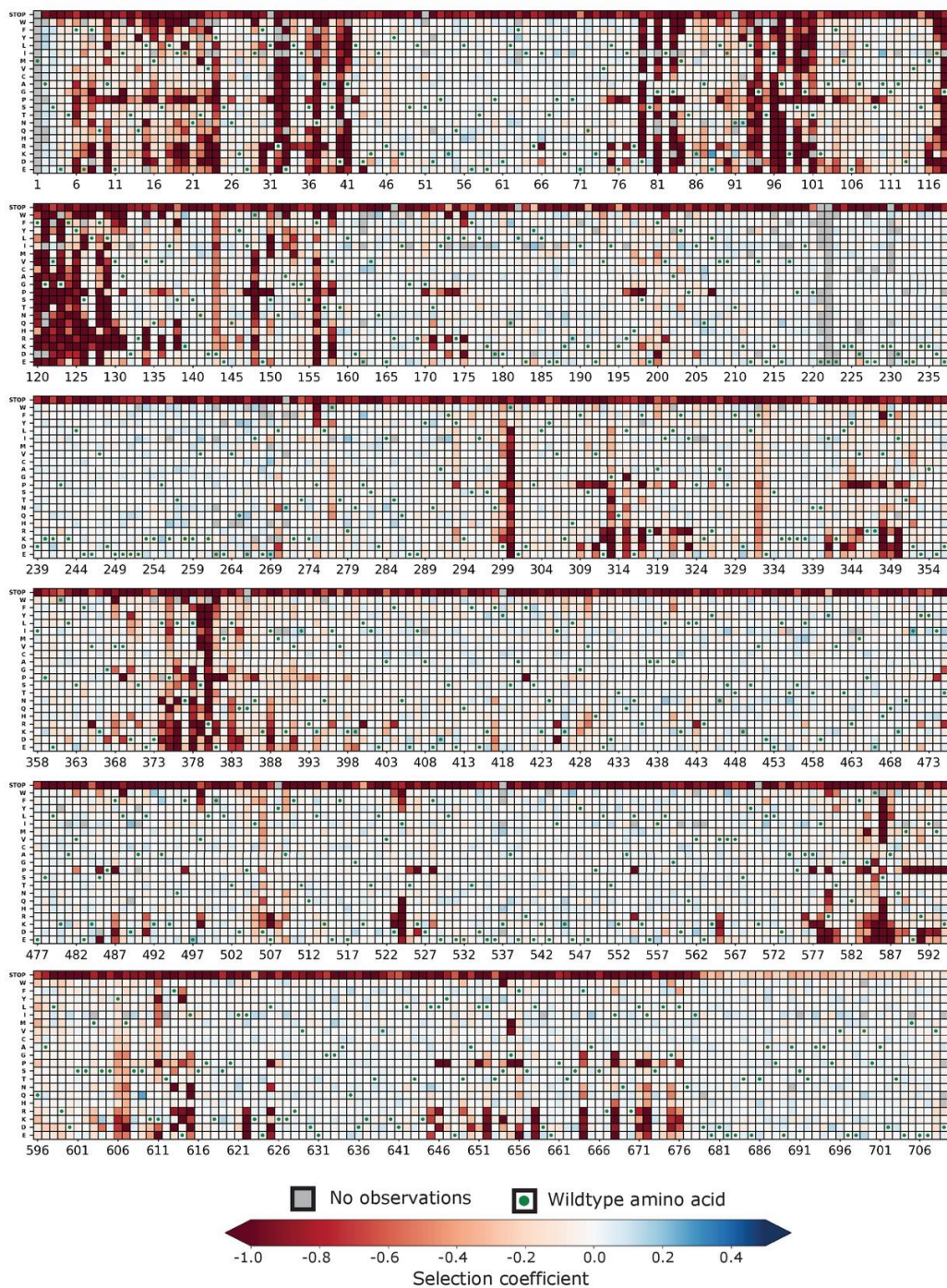
**Supplementary Figure 3.2: Measurement of selection coefficients for positions 2–220 in this study correlated strongly ( $R^2 = 0.87$ ) with estimates of the Hsp90 N-domain in a previous study (Mishra et al., 2016), indicating that biological replicates show high reproducibility. The blue dashed line indicates the line of best fit.**



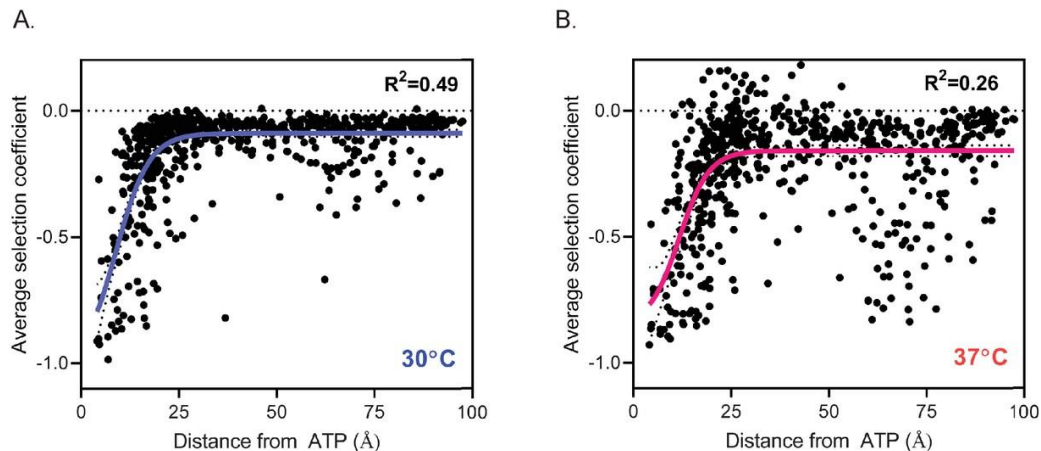
### Supplementary Figure 3.3: Analysis of variation in stop codon selection

**coefficients (A)** The average initial reads measured per codon for each amino acid. **(B)** The difference in the selection coefficients between stops in the two standard replicates compared to the initial reads for the corresponding stop in replicate 1. The blue dashed line indicates the line of best fit.  $R^2 = 0.022$ ,  $p < 0.0001$ . **(C)** The distribution of selection coefficients for each stop codon in standard replicate 1.

## Standard Replicate 1 (30°C)

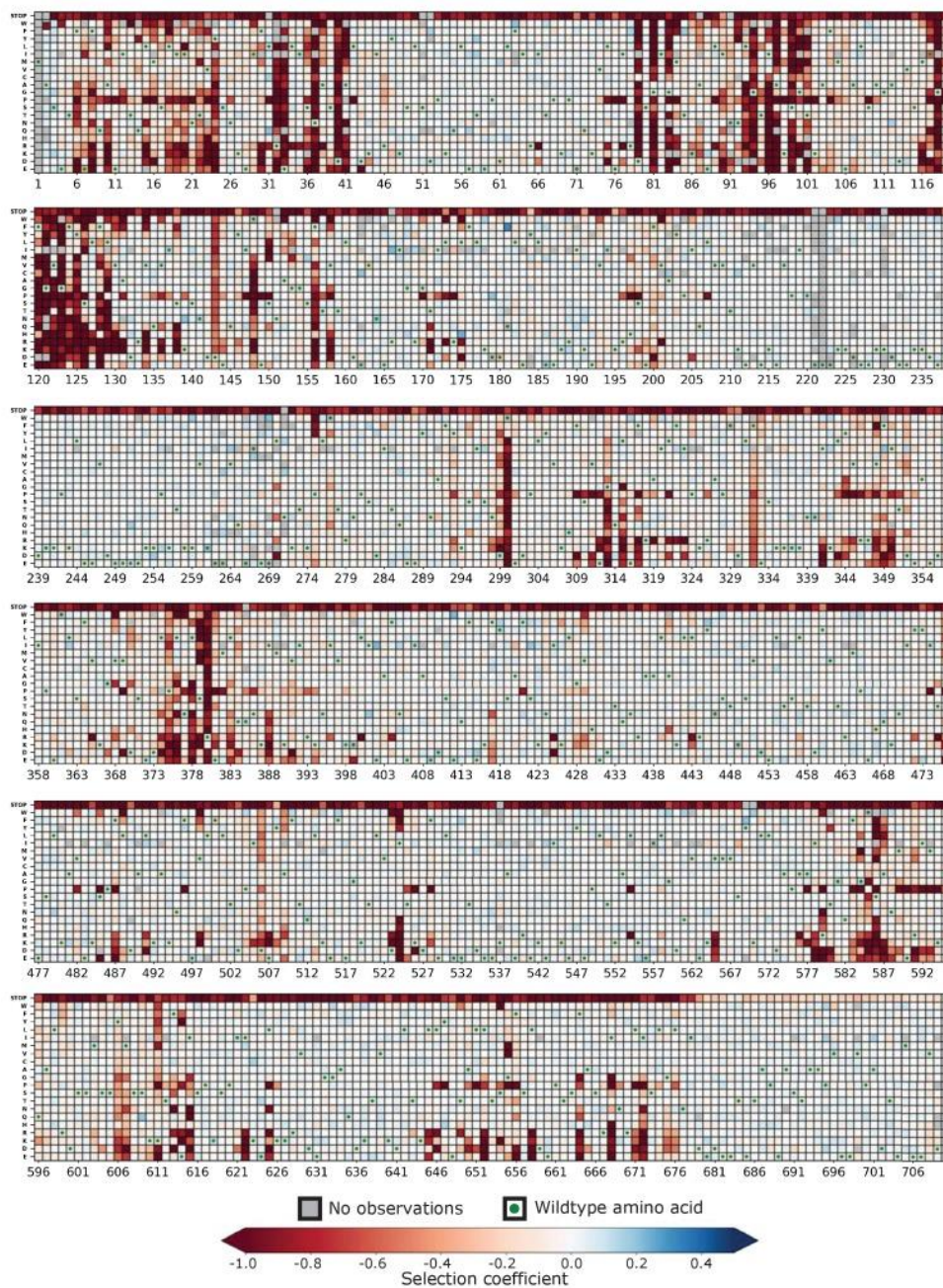


**Supplementary Figure 3.4: Heatmap representation of the selection coefficients observed for single amino acid changes across amino acids 2–709 of Hsp90 in standard (30°C) conditions in replicate 1.**



**Supplementary Figure 3.5: Correlation of mutational sensitivity with distance to ATP.** (A) The average selection coefficient at each position in standard (30°C) conditions correlates with distance to ATP ( $R^2 = 0.49$ ). (B) The average selection coefficient at each position at 37°C does not correlate as well with distance to ATP as in standard conditions ( $R^2 = 0.26$ ). The data points were fit to a sigmoid function.

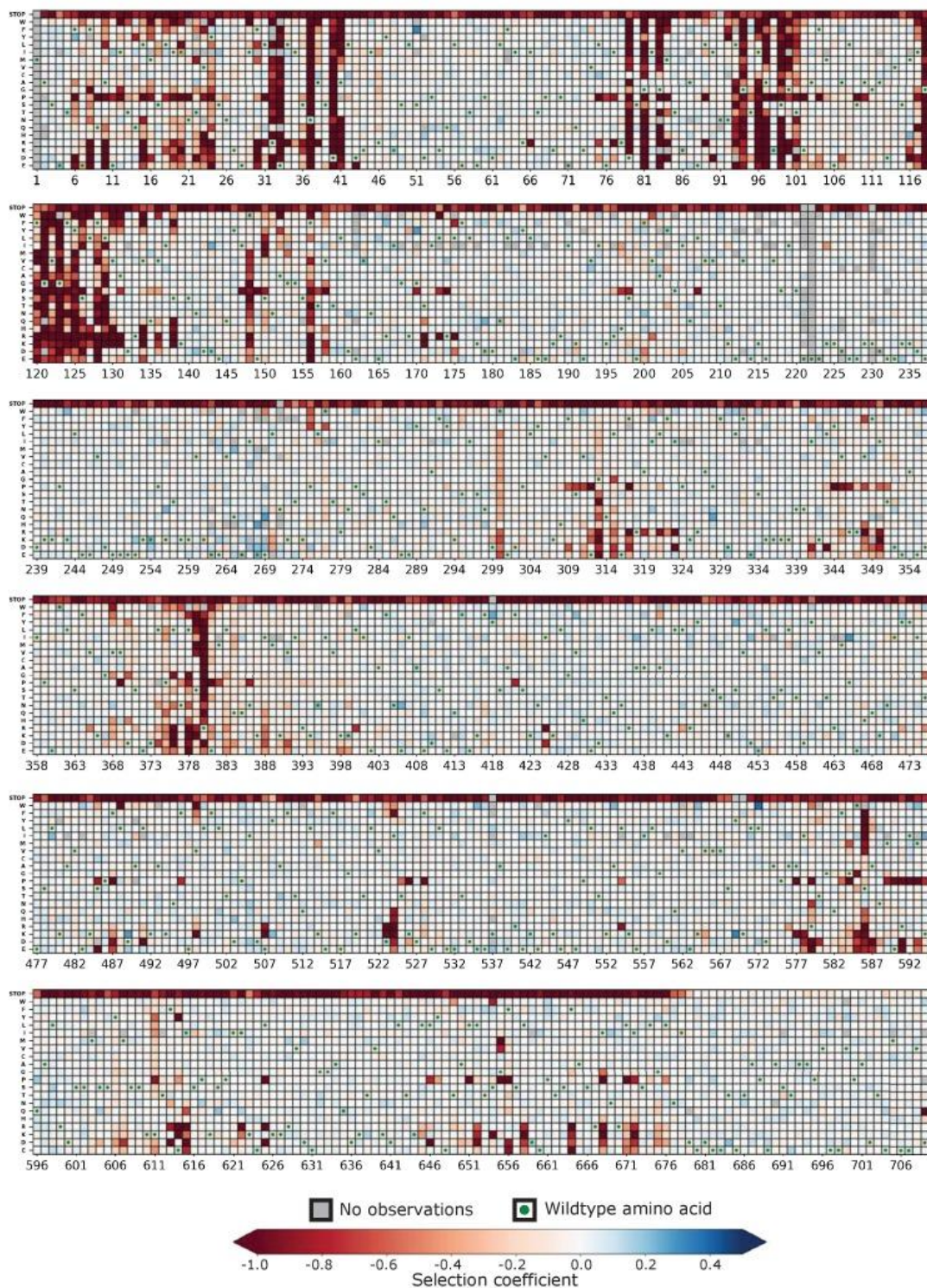
## Nitrogen Depletion



**Supplementary Figure 3.6: Heatmap representation of the fitness map observed for single amino acid changes of Hsp90 in nitrogen depletion.**

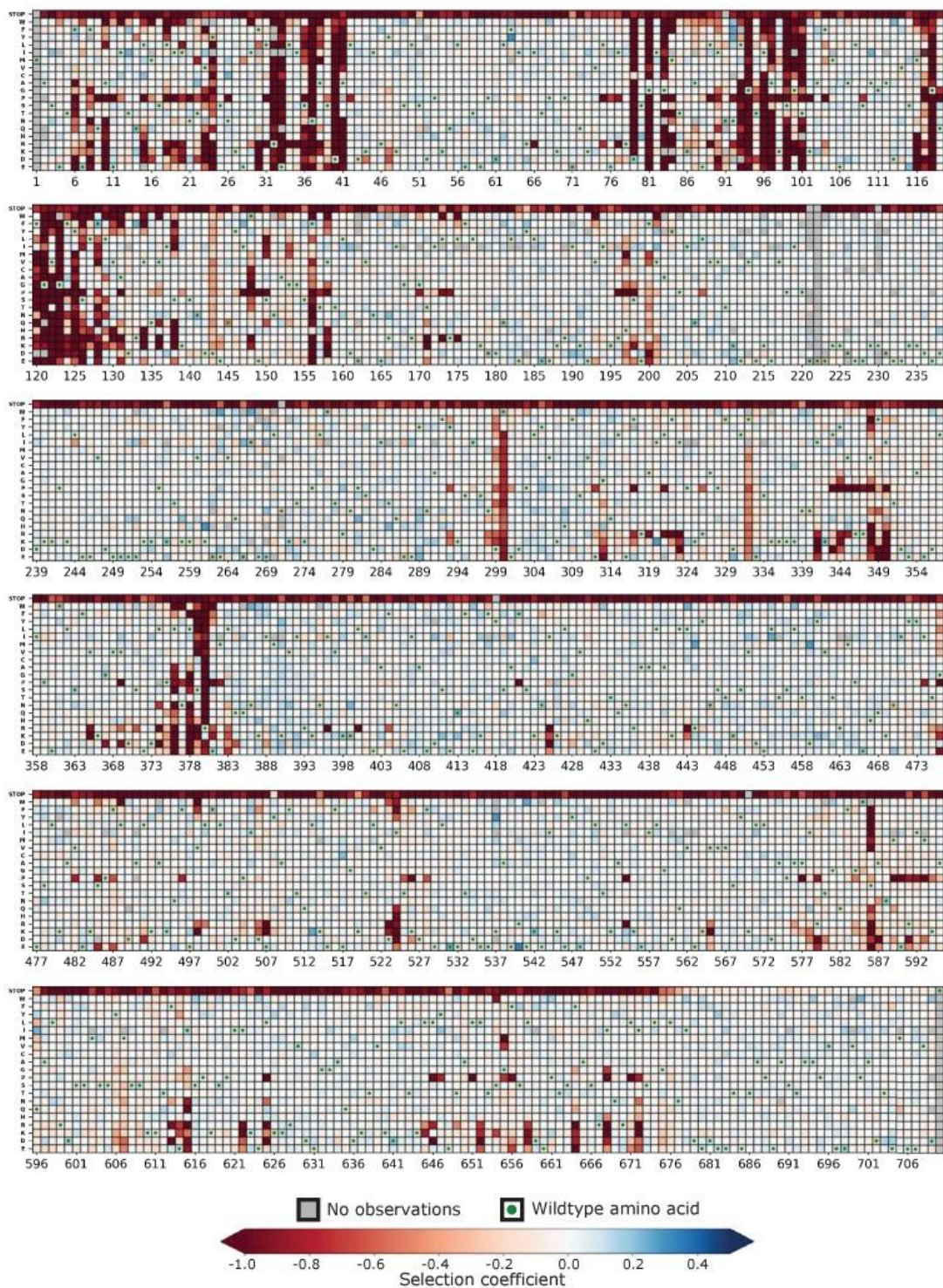


## Salt



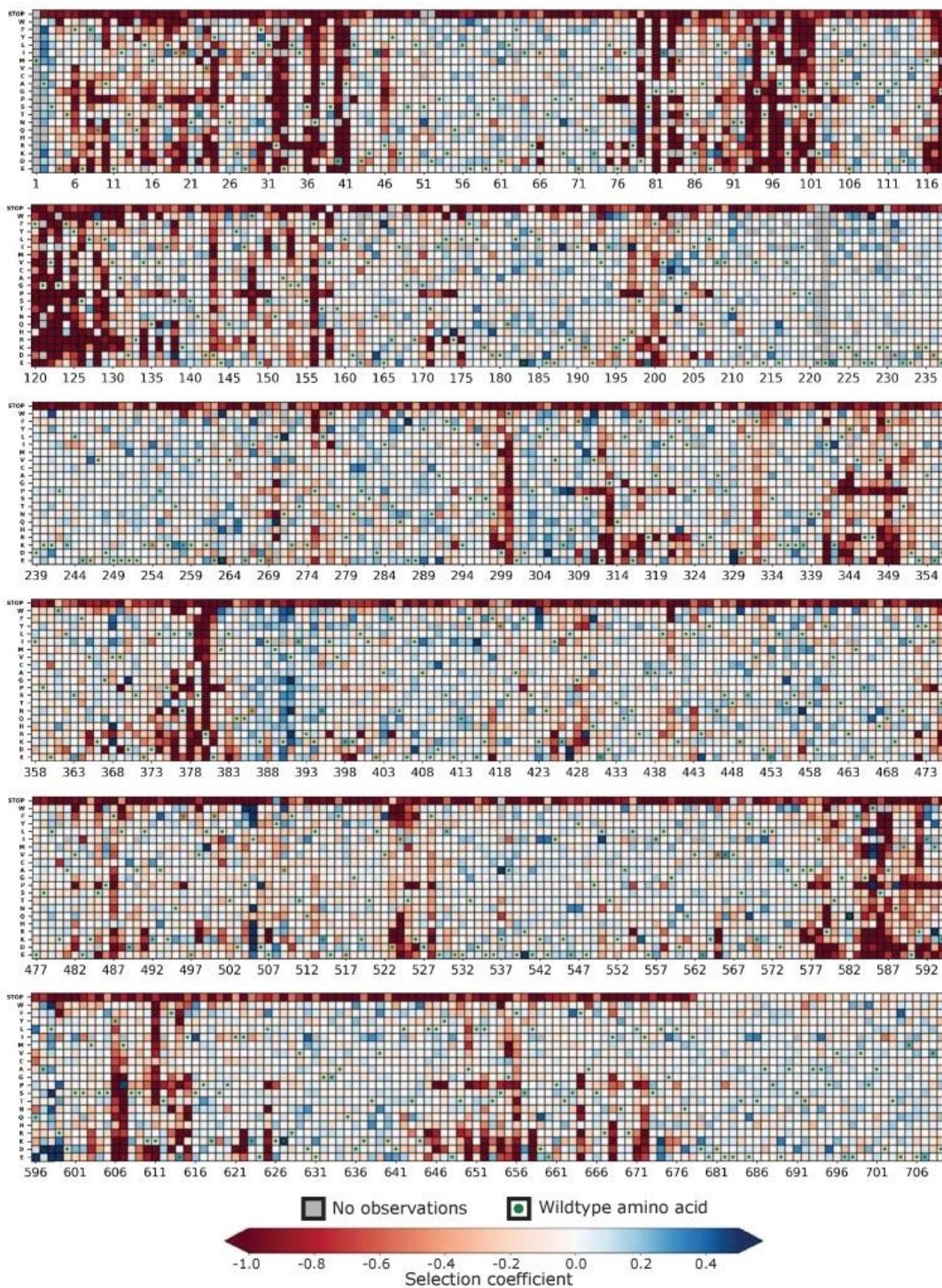
**Supplementary Figure 3.7: Heatmap representation of the fitness map observed for single amino acid changes of Hsp90 in salt.**

## Ethanol



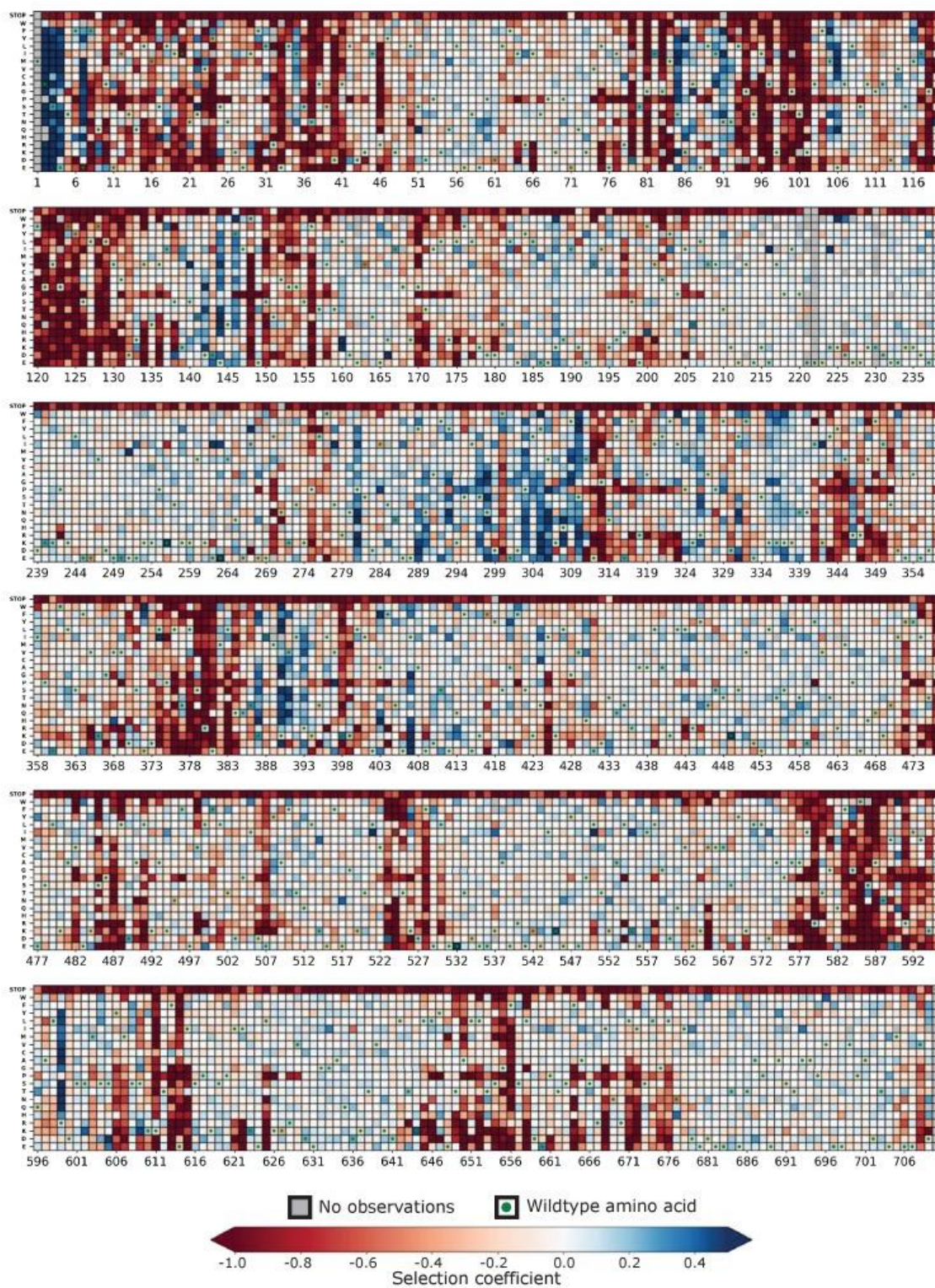
**Supplementary Figure 3.8: Heatmap representation of the fitness map observed for single amino acid changes of Hsp90 in ethanol.**

## Diamide

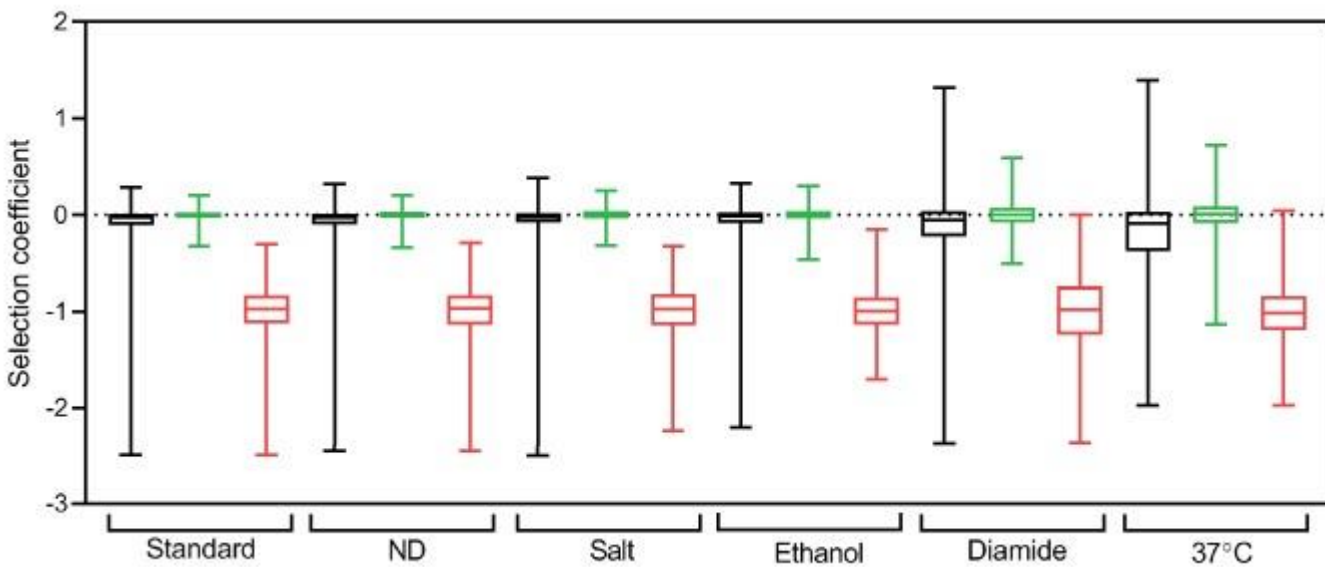


**Supplementary Figure 3.9: Heatmap representation of the fitness map observed for single amino acid changes of Hsp90 in diamide.**

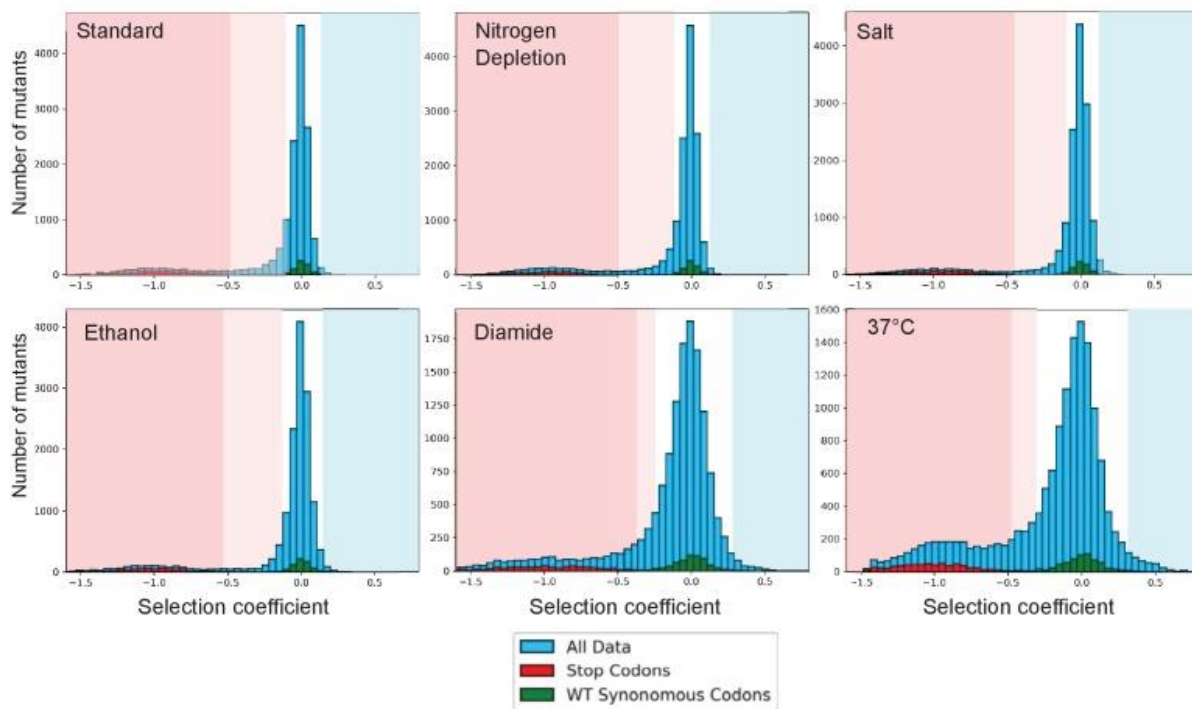
## High temperature (37°C)



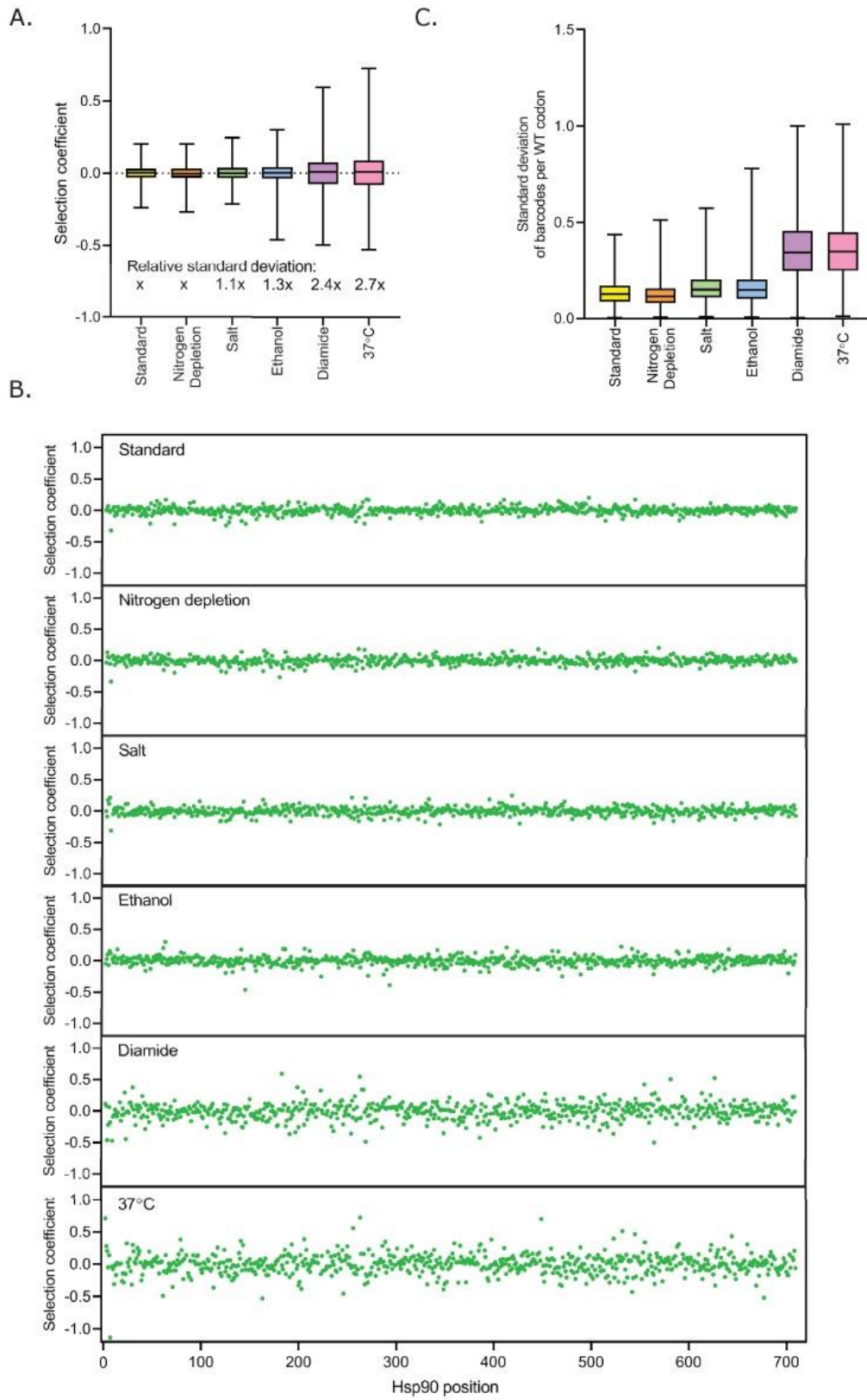
**Supplementary Figure 3.10: Heatmap representation of the fitness map observed for single amino acid changes of Hsp90 at 37C.**



**Supplementary Figure 3.11: Distribution of selection coefficients for non-synonymous mutations (black), wild-type synonyms (green), and stops (red) in each environmental condition.**

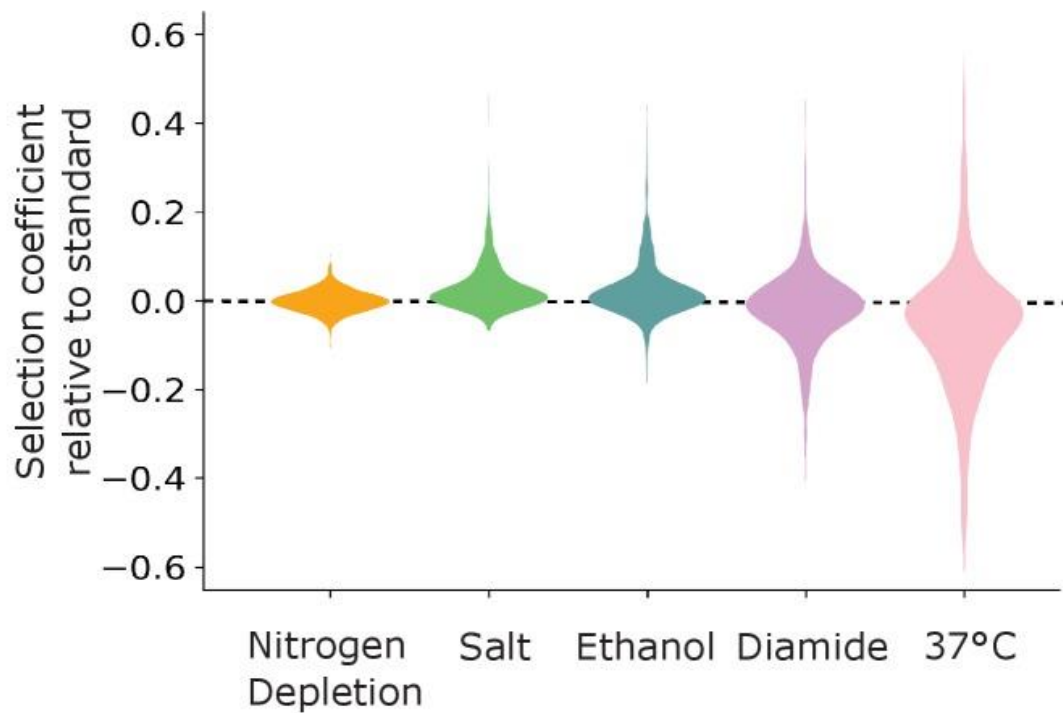


**Supplementary Figure 3.12: Distribution of selection coefficients in each environmental condition.** Mutations were categorized as beneficial (light blue shading), wild-type-like (white shading), intermediate (light pink shading) or deleterious (dark pink shading) based on the distribution of wild-type synonyms and stop codons in each condition.

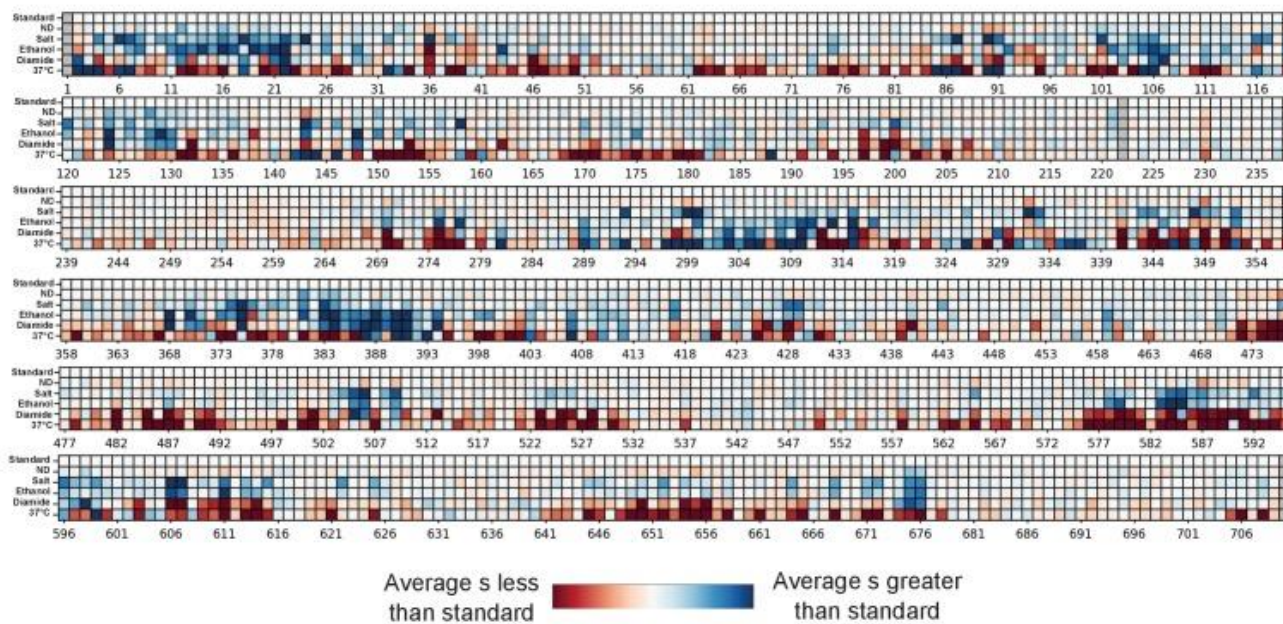


**Supplementary Figure 3.13: Analysis of variation in wild-type synonym selection coefficients.** (A) Distribution of selection coefficients for the wild-type synonyms in each condition. The standard of deviation of wild-type synonyms in each condition relative to standard conditions ( $x$ ) is specified under the corresponding box. (B) The selection coefficient for wild-type synonyms at each position of Hsp90 in each condition. (C) The variation of selection coefficients for barcodes with greater than 50 initial reads for each wild-type codon in each condition.

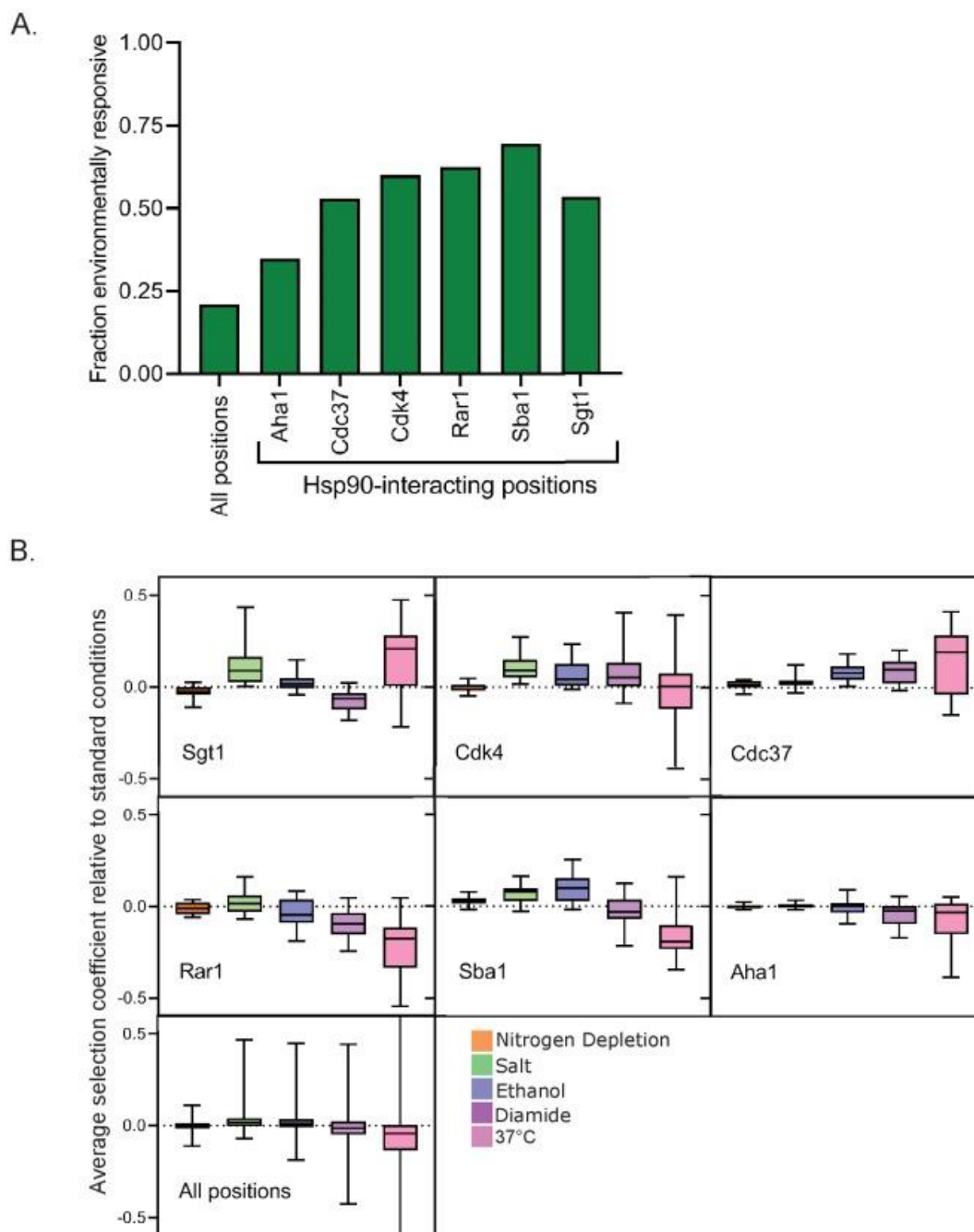




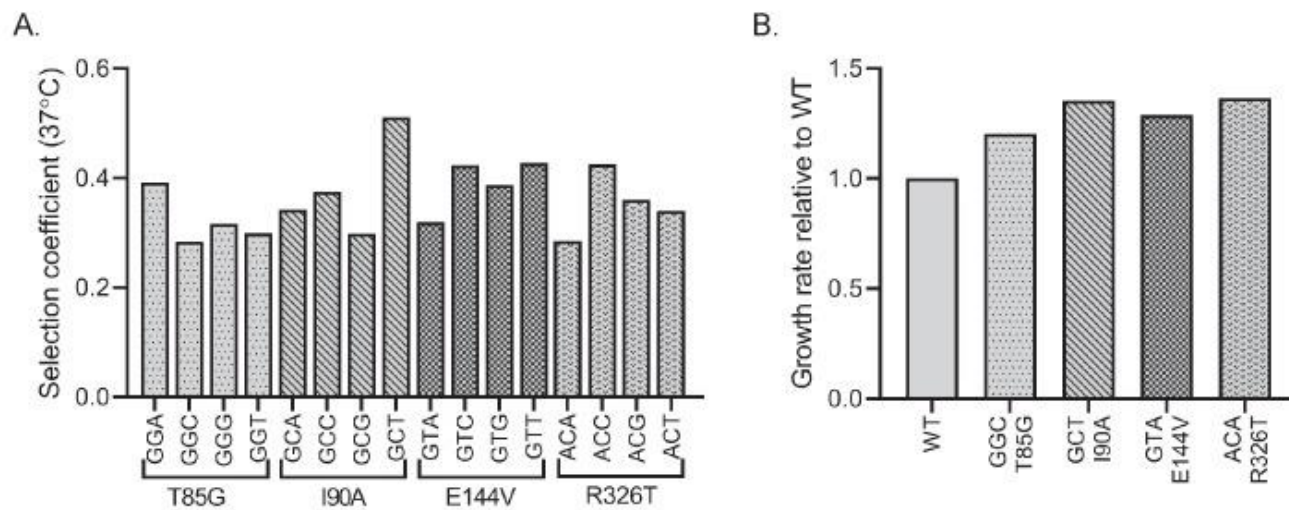
**Supplementary Figure 3.14: Distribution of the difference between selection coefficients of each mutation in each stress condition and the same mutation in standard conditions.**



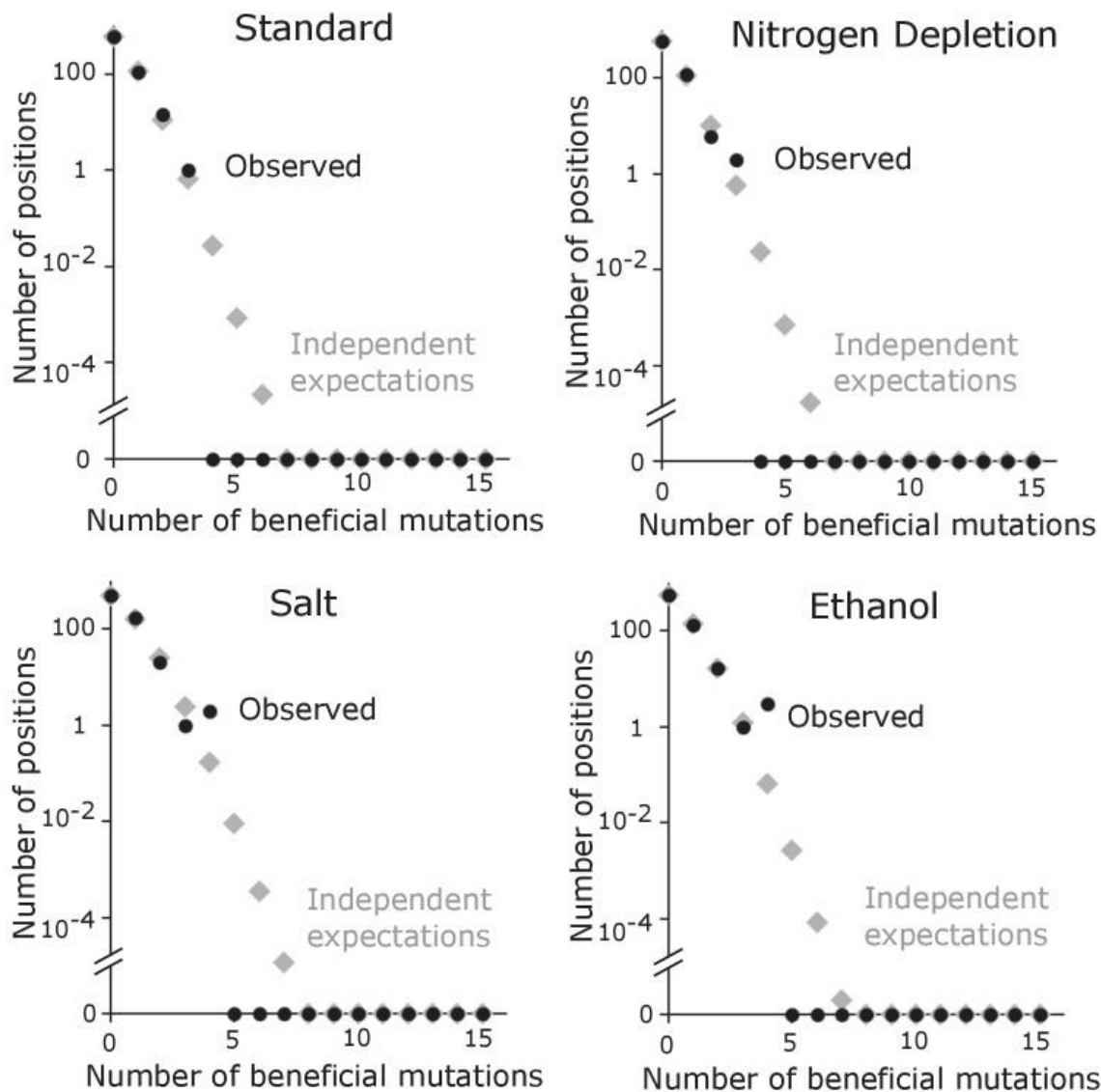
**Supplementary Figure 3.15: Heatmap representation of the average selection coefficient ( $s$ ) at each position in each environmental condition relative to the average selection coefficient at the same position in standard conditions.**



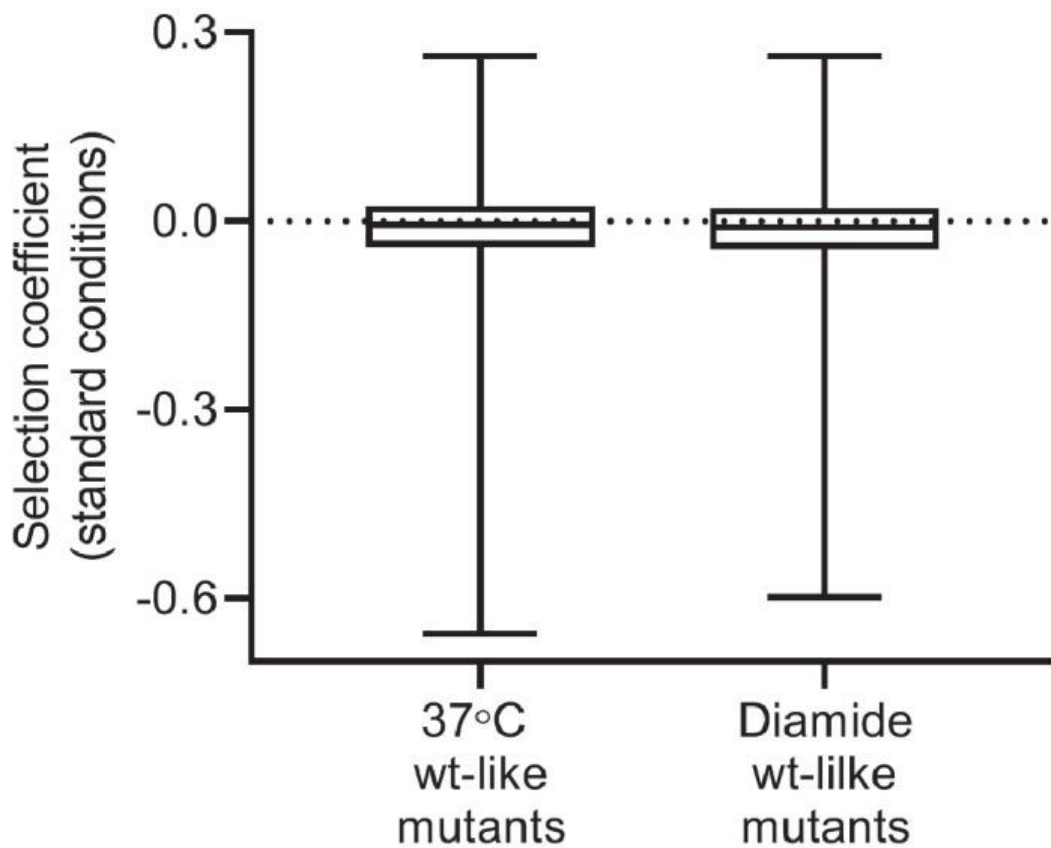
**Supplementary Figure 3.16: Environmentally responsive Hsp90 positions are enriched in binding contacts.** (A) The fraction of Hsp90 positions at interfaces that were categorized as environmentally responsive. (B) The average selection coefficient in each environment relative to standard at all the Hsp90 positions at each stated interface



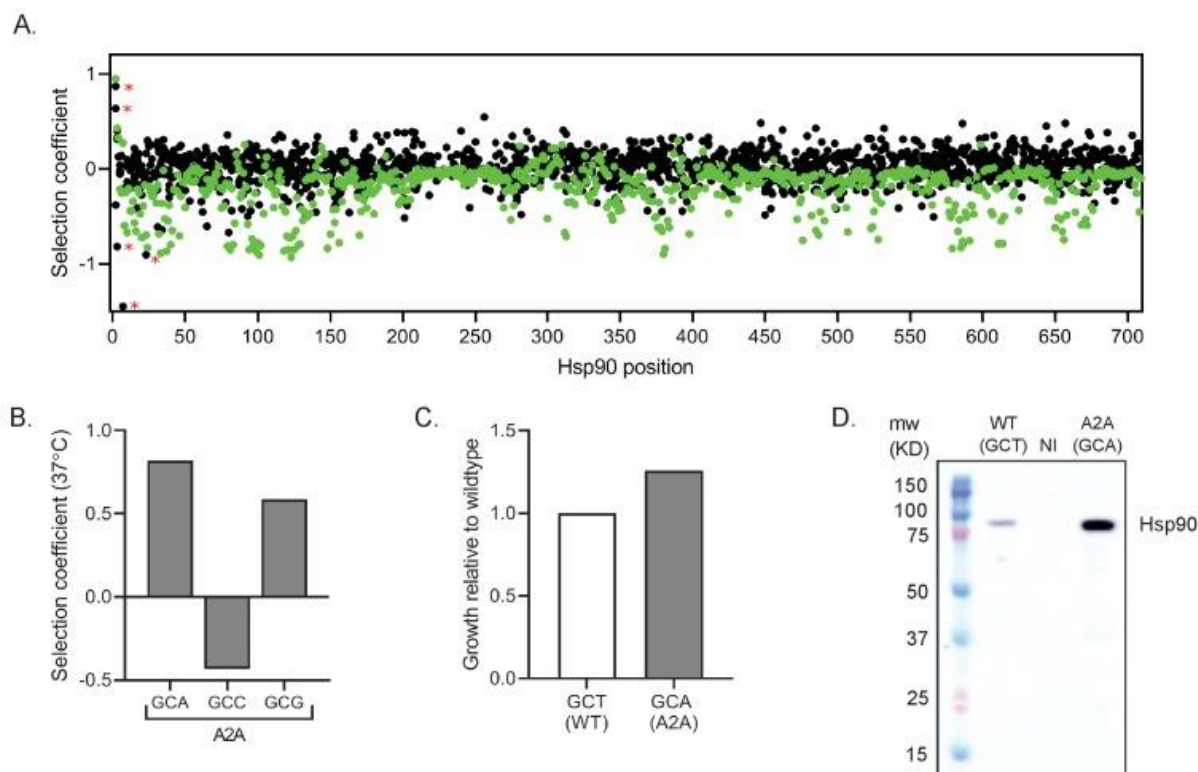
**Supplementary Figure 3.17: Validation of beneficial mutants at 37°C.** (A) Selection coefficients of synonymous codon variants for four amino acid mutants with beneficial selection coefficients at 37°C show high correlation. (B) The same individual variants analyzed in isolation exhibit increased growth rates.



**Supplementary Figure 3.18: Distribution of the number of beneficial mutations at the same position in standard, nitrogen depletion, salt, and ethanol conditions.** Independent expectations were calculated as the probability of the stated number of mutations occurring at the same position by chance.

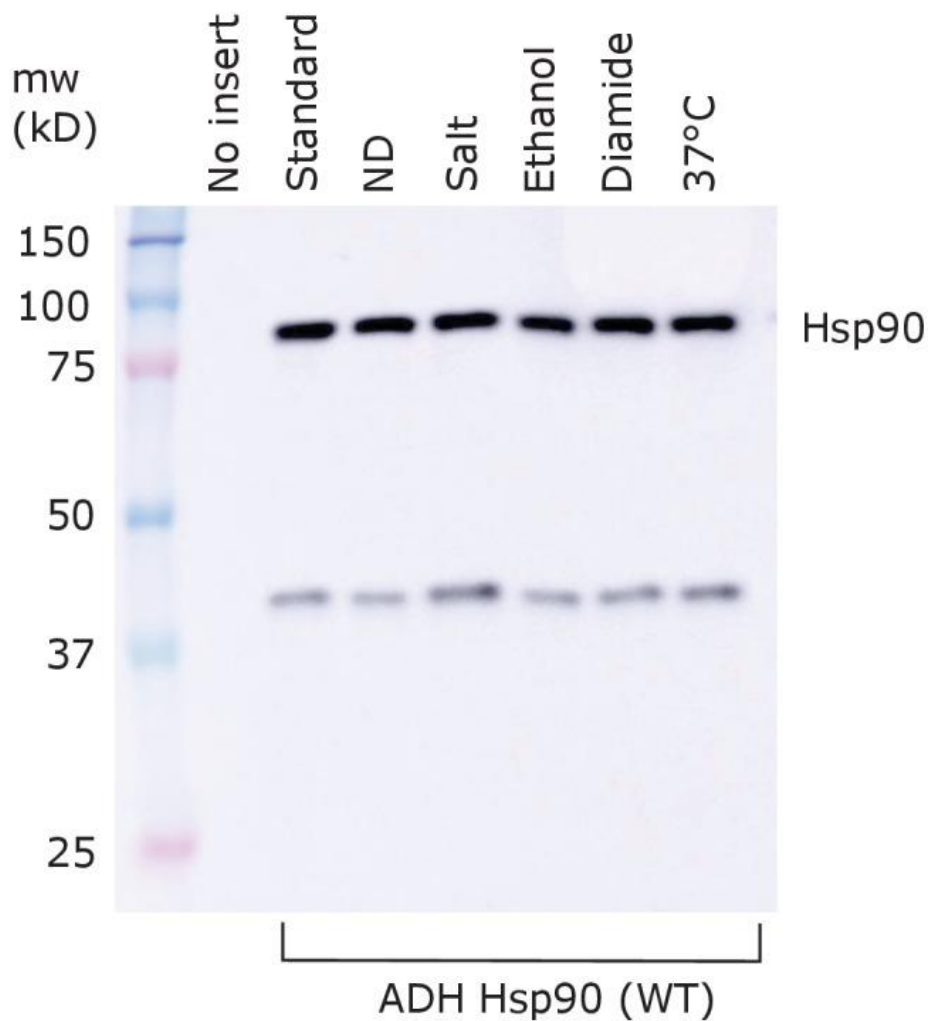


**Supplementary Figure 3.19: Selection coefficients in standard conditions for all wild-type-like mutations at 37°C and in diamide.**



### Supplementary Figure 3.20: Synonymous mutations at the beginning of Hsp90 have strong beneficial growth effects.

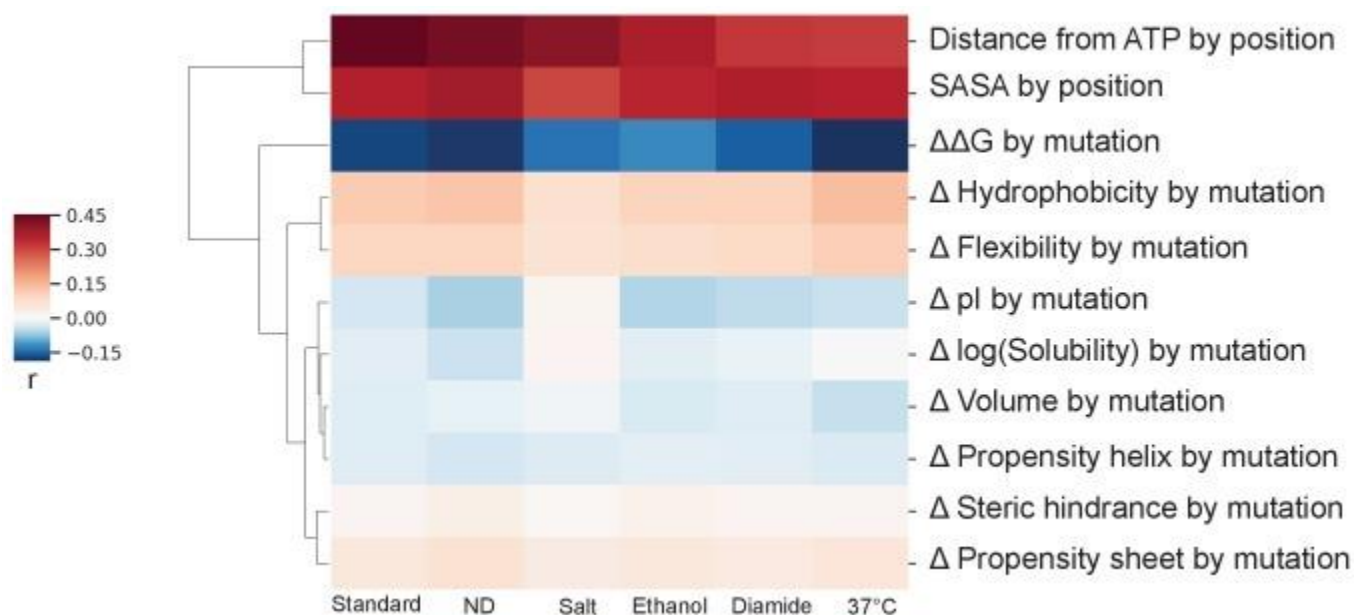
(A) Selection coefficients for all codon variants of synonymous mutations at each position of Hsp90 (black) at 37°C compared to the average selection coefficient of all mutations at each position (green) at 37°C. Synonymous mutations that were deemed beneficial or deleterious after a Bonferroni correction are noted with a red asterisk. (B) Selection coefficients for codon variants of synonymous mutants at position Ala2 show high variation. (C) The individual Hsp90 variant containing the synonymous codon mutation GCT to GCA at position 2 exhibited an increased growth rate 37°C. (D) The individual synonymous mutant variant GCA at position two exhibited higher cellular expression levels at 37°C. WT: wild-type Hsp90 (GCT at 2<sup>nd</sup> position), NI: No insert, A2A: Hsp90 with the GCT to GCA codon mutation.



**Supplementary Figure 3.21: Hsp90 expression in the Hsp90 shutoff yeast strain harboring either wild-type Hsp90 under the constitutive ADH promoter or a null plasmid (no insert).**

Cells were grown for 12 hours in dextrose media at 30°C and then were transferred into the individual environmental conditions and grown for eight additional hours. Hsp90 levels were monitored by western blotting.



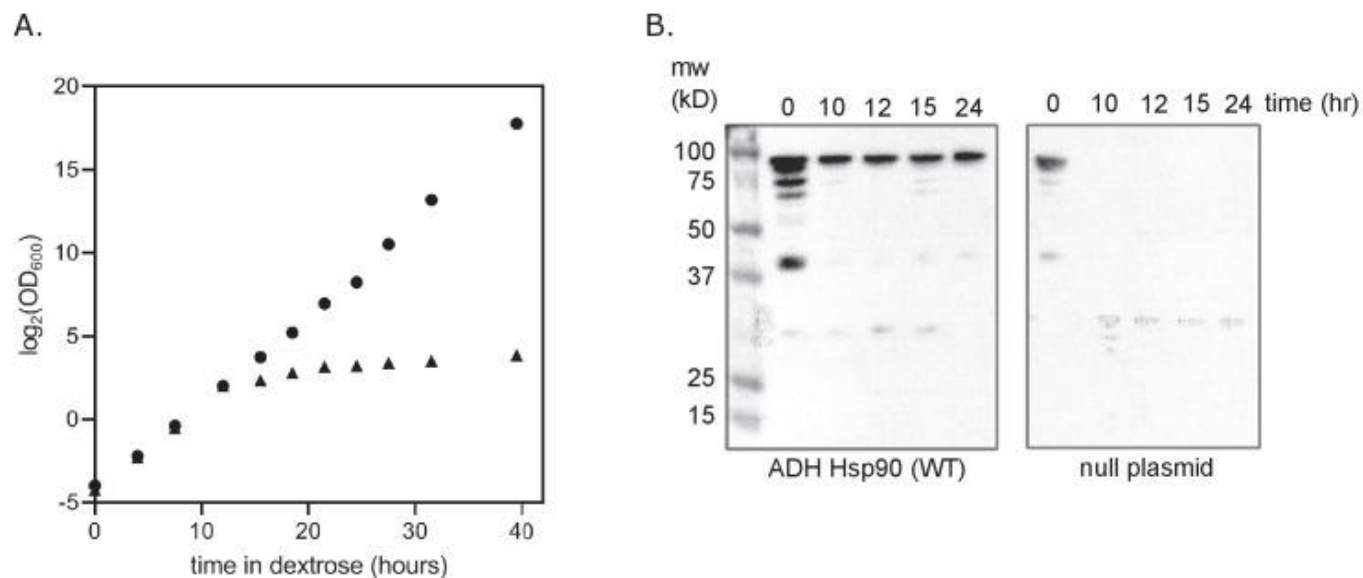


**Supplementary Figure 3.22: Heatmap of Spearman's rank correlation coefficients ( $r$ ) between molecular features (rows) and selection coefficients per mutation or per position (average selection coefficient for all amino acids at the position) for each environment (columns).** Figure 6—figure supplement 2—source data 1

The Spearman's rank correlation coefficients ( $r$ ) and associated p-values between molecular features and selection coefficients for each environment.

<https://cdn.elifesciences.org/articles/53810/elifesciences-53810-fig6-figsupp2-data1-v2.docx>

[Download elifesciences-53810-fig6-figsupp2-data1-v2.docx](#)



**Supplementary Figure 3.23: Validation of yeast Hsp90-shutoff strain.**

(A) Growth of Hsp90 shutoff yeast harboring either wild-type (WT) Hsp90 under the ADH promoter (●) or a null plasmid (▲). Growth was monitored by optical density (OD) at 600 nm. (B) Hsp90 expression in cells harboring the same plasmids grown in dextrose for the stated number of hours. Hsp90 protein levels were monitored by western blotting.

**Table 3.1****Key resources table**

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Gene ( <i>Saccharomyces cerevisiae</i> )	<i>hsp82</i>	<i>Saccharomyces</i> Genome Database	SGD: S000006161	Hsp90 chaperone
Antibody	anti-Hsp90 $\alpha/\beta$ (Mouse monoclonal)	Cayman chemical; Cat# 10011439	RRID:AB_10349777	WB (1:3000)
Recombinant DNA reagent	Barcoded Hsp90 plasmid library	This paper		See Materials and methods for library construction
Recombinant DNA reagent	p414ADH $\Delta$ ter plasmid	PMID: 23825969		
Recombinant DNA reagent	p414GPD plasmid	PMID: 23825969		
Commercial assay or kit	BCA Protein Assay Kit	Pierce	Cat #23227	
Commercial assay or kit	KAPA SYBR FAST qPCR Master Mix	Kapa Biosystems	KK4600	
Chemical compound, drug	Diamide	Sigma Aldrich	D3648	
Software, algorithm	Barcode – Hsp90 ORF assembly	This paper	<a href="https://github.com/JuliaFlynn/Barcode_ORF_assembly">https://github.com/JuliaFlynn/Barcode_ORF_assembly</a>	Associates barcodes with open reading frame mutations from paired end sequencing data (Flynn, 2020a; copy archived at <a href="https://github.com/elifesciences-publications/Barcode_ORF_assembly">https://github.com/elifesciences-publications/Barcode_ORF_assembly</a> )
Software, algorithm	Tabulate Hsp90 counts	This paper	<a href="https://github.com/JuliaFlynn/Tabulate_counts">https://github.com/JuliaFlynn/Tabulate_counts</a>	Counts Hsp90 alleles from raw fastq files and barcode_orf assembly file (Flynn, 2020b; copy archived at <a href="https://github.com/JuliaFlynn/Tabulate_counts">https://github.com/JuliaFlynn/Tabulate_counts</a> )
Software, algorithm	EmpiricIST	PMID: 30127529	<a href="https://github.com/Matu2083/empiricIST">https://github.com/Matu2083/empiricIST</a>	Estimates selection coefficients based on the MCMC approach
Sequenced-	Sequencing	This paper		See <i>Supplementary file 1</i>
Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Sequenced-based reagent	Site-directed mutagenesis primers	This paper		See <i>Supplementary file 1</i>
Sequenced-based reagent	Library construction oligomers	This paper		Available upon request

## Chapter IV: General Discussion

### Summary

Environments in nature are constantly fluctuating, thus presenting selective pressures that can impact an individual's fitness at many biological levels, including the sequence level. Therefore, environmental challenges can alter the DFE of new mutations and have direct evolutionary consequences. As a result, the interactions between genes and environments may contribute to molecular evolution and adaptation, however very few studies have accurately quantified the impact of environment on the DFE of new mutations (Dhar et al. 2011; Hietpas et al. 2013; Arribas et al. 2014; Mavor et al. 2016; Dandage et al. 2017; Li and Zhang 2018), especially in regards to adaptive mutations and their costs in alternate environments. The basis of the work presented within this dissertation is focused on accurately measuring the impact of novel environmental challenges on the DFE of new mutations within the yeast, *S. cerevisiae* Hsp90 sequence to identify environment dependent mutations, quantify their adaptive potential and costs. Furthermore, I seek to understand how selection pressure on the Hsp90 protein sequence and function changes upon exposure to different environmental challenges and can facilitate yeast adaptation to novel environmental stress. Yeast Hsp90 was chosen for the focus of this dissertation because yeast Hsp90 has important roles in both short-term and long-term adaptation. Specifically, Hsp90 is required to buffer proteotoxic and deleterious fitness effects that arise during short-term environmental stress and can disrupt protein and cellular function, ultimately affecting organismal fitness and survival (Gasch et al. 2000; Yang et

al. 2006; Richter et al. 2010). Additionally, the long-term buffering effects of Hsp90 can compromise Hsp90's role in maintenance of standing genetic variation, resulting in the emergence and evolution of new traits that facilitate rapid genetic adaptation to new stress conditions (Rutherford and Lindquist 1998; Rutherford 2003; Jarosz and Lindquist 2010).

In Chapter II, I present a modified version of the EMPRIC approach to systematically quantify how the costs of adaptation vary for a large region of Hsp90 by studying the environmental impact of evolutionary relevant variables including, optimal growth conditions, and conditions that affect yeast growth rate and have been previously identified to impact the upregulation of Hsp90 expression including, osmotic (salinity and sorbitol) and oxidative (H<sub>2</sub>O<sub>2</sub> and Diamide) stress (Gasch et al. 2000) on the DFE of, ~ 2300 amino acid changing mutations for a 119 amino acid region of the middle domain of yeast Hsp90, (amino acids 291-409) at normal expression. This 119 amino acid region was chosen for this study because it is highly conserved and has been previously shown to be involved in Hsp90 client-binding and client maturation function, stabilization of the Hsp90 dimer, stabilization of the N-terminal-Middle and Middle-C terminal domain and regulation of ATPase-Chaperone activity (Nathan and Lindquist 1995; Nathan et al. 1997; Meyer et al. 2003; Hawle et al. 2006; Hagn et al. 2011). As a result, mutations within this region could impact diverse aspects of Hsp90 such as, relative affinity and priority of different clients involved in different stress response pathways, which may provide an adaptive benefit to specific environments. We discover that the fitness effects of mutations within this region of Hsp90 correlates well across environments, with diamide standing out with respect to mutant fitness

effects. These general results correlate with previous studies that have reported that mutations had similar fitness effects across environments that shared metabolic features (Ostrowski et al. 2005; Dillon et al. 2016; Sane et al. 2018). We identify adaptive mutations under all conditions with little cost of adaption under alternate environments, with the exception of diamide which showed the largest proportion and magnitude of beneficial and costly mutations. We identify protein regions that are enriched in beneficial, deleterious and costly mutations that coincides with residues involved in stabilization of Hsp90 client-binding interfaces, stabilization of Hsp90 interdomains and ATPase chaperone activity. Lastly, we find that the diversity of natural amino acid variants in observed Hsp90 middle domain sequences of extant eukaryotes supports robust growth under all conditions, consistent with our observations and indicating that fluctuating environments may place long-term selective pressure on Hsp90 that results in robustness under diverse conditions. Together these results provide information regarding the role and adaptive potential of the middle domain of Hsp90 that complements and extends previous knowledge.

In chapter III, I use the modified version of the EMPIRIC approach to quantify the growth effects or “fitness” of new mutations within the entire yeast Hsp90 sequence ~ 44,604 single codon changes encoding 14,160 amino acid variants under optimal growth conditions and environments that partially decrease the growth rate of yeast including: temperature shock, nitrogen deprivation, ethanol stress, oxidative stress, and osmotic stress under low Hsp90 expression. We present comprehensive fitness maps of an entire protein under each condition and compare to the historical record of accrued Hsp90 mutations within 261 extant eukaryotes. We find that each condition had

distinct impacts on the growth of many variants, indicating that environment can have a large impact on the evolution of Hsp90. We found that environmentally sensitive residues coincide with client-co-chaperone interaction sites and interdomains, suggesting that mutations within Hsp90 may impact client-binding- co-chaperone interactions, structural dynamics and stability that may facilitate changes in relative affinity and or priority of clients that restructures stress response networks. We identified adaptive mutations under individual conditions that correlated with costs of adaptation in alternate environments. Similar to chapter II, we found the largest proportion and magnitude of adaptive and costly mutations under diamide conditions in addition to heat that coincides with functional hotspots. However, we find that natural variants of Hsp90 support growth in all environments, providing further evidence that suggests that selection for robust growth to diverse stress conditions has shaped the natural evolution of Hsp90.

### **Most Mutations Exhibit Neutral and Deleterious Fitness Effects Across Environments**

In both studies we observe DFEs with mostly wild-type like mutations under all conditions with the highest proportion of neutral mutations observed under standard conditions. These results are consistent with observations of large numbers of wild-type like mutations in previous mutational scanning analysis of DFEs (Soskine and Tawfik 2010; Hietpas et al. 2013; Melamed et al. 2013; Bank et al. 2014; Hom et al. 2019), enrichment of wild-type like mutations in the subset of naturally observed amino acid variants across 261 eukaryotic sequences examined and predictions by Ohta and Kimura on the bimodal distribution of mostly neutral and deleterious mutant fitness

effects and in the context of the near neutral model of evolution (Ohta 1973; Kimura 1983; Ohta 1992). These results though coinciding with previous studies and predictions were at odds with the strong conservation of Hsp90 across large evolutionary distances. There are many biological and experimental factors that could potentially explain why we observed these results including: selection pressures in the laboratory might differ from those in nature, whereby natural environments are different and rapidly fluctuating instead of stagnant (Mustonen and Lässig 2009), secondly there are differences between experimental resolution and the resolution at which natural selection acts upon in nature (Reznick and Ghalambor 2005; Kvitek and Sherlock 2013). Specifically, the resolution of our experiment is most likely to be lower than the resolution of large yeast populations in natural environments, whereby experimentally unmeasurable fitness defects could be strongly subjected to purifying selection in nature over long evolutionary time-scales. Thirdly, because our studies were done using haploid yeast that reproduce asexually via mitosis it does not mimic what would occur in nature, whereby wild type yeast can occur as haploid cells that can mate with other mating type haploid cells to produce stable diploid cells that can withstand environmental insults (Haber 2012). Furthermore, diploid yeast cells can undergo sporulation in response to environmental stress, such as nitrogen starvation (Freese et al. 1982). Finally, because Hsp90 has been shown play a direct role in the vegetative growth, reproduction, and virulence of the ascomycete fungus *Fusarium graminearum* (Bui et al. 2016) and the filamentous fungal pathogen *Aspergillus fumigatus* (Lamoth et al. 2012), mutations to Hsp90 may have varying effects on experimental asexual yeast populations than what we would observe in natural yeast populations. Together these



factors could have impacts on the DFEs of mutations in Hsp90 across environments and the evolutionary consequences observed here.

We also identified a large number of deleterious mutations across all environments, consistent with previous predictions by Ohta and Kimura that most mutations are deleterious (Ohta 1973; Kimura 1983; Ohta 1992) and previous observations of large proportions of deleterious mutations in studies of DFE of mutations (Wloch et al. 2001; Sanjuán et al. 2004; Hietpas et al. 2011; Hietpas et al. 2012; Hietpas et al. 2013; Bank et al. 2014). We identified the largest proportion of deleterious mutations under diamide and heat stress consistent with the observation that both conditions negatively affect growth rate more than other environments. Mapping deleterious mutations to Hsp90 structure in Chapter II revealed residues that coincides with regions towards the end of the beneficial hotspot that partly overlap with the catalytic loop required for ATPase chaperone activity, a residue located in the ATPase-Middle domain interface, and a residue involved in client-binding. In chapter III we find that strongly deleterious mutations coincide with regions of Hsp90 important for Hsp90 stability, client binding and ATPase chaperone activity. These results are consistent with previous observations of deleterious mutations within these regions impacting function and growth or fitness (Meyer et al. 2003; Mishra et al. 2016; Wolmarans et al. 2016). Together these results indicate that these regions are under strong purifying selection. The strong purifying selection for disadvantageous mutations within these regions of Hsp90 seems logical because residues within these regions are important for Hsp90 function and cell viability (Meyer et al. 2003) and ATP hydrolysis is necessary for the activation of all clients (Wolmarans et al. 2016).

## **The Impact of Stress on Average Mutant Fitness Effects and Selection of Hsp90 Mutations**

In Chapter III, we found that stress conditions had distinct impacts on average mutant fitness effects and selection of Hsp90 mutations when compared to standard conditions. Specifically, ethanol and salt had less deleterious effects on mutant fitness, resulting in decreased strength of selection on Hsp90 mutations. These results indicate decreased demand for Hsp90 function and agree with the observation that less Hsp90 has a minimal impact on yeast growth under these conditions, whereas, diamide and heat resulted in stronger selection of Hsp90 mutations that lead to a greater number and magnitude of deleterious mutations. These findings overlap with the theory of metabolic flux (Kacser and Burns 1981; Dykhuizen et al. 1987), whereby gene products that are rate limiting for growth are subject to stronger selection. Furthermore, these results are consistent with the hypothesis that heat and diamide stress may cause a growth limiting increase in unfolded Hsp90 clients that is rate limiting for growth and thus requires more Hsp90 function, which is compromised under these conditions. When mapping environmentally sensitive positions to structure we find that these environmentally responsive positions were located throughout all domains of Hsp90, and enriched in client and co-chaperone interacting sites and intramolecular Hsp90 contacts involved in transient conformational changes (Bohen and Yamamoto 1993; Nathan and Lindquist 1995; Meyer et al. 2003; Meyer et al. 2004; Roe et al. 2004; Ali et al. 2006; Hawle et al. 2006; Retzlaff et al. 2009; Zhang et al. 2010; Hagn et al. 2011; Genest et al. 2013; Lorenz et al. 2014; Verba et al. 2016; Kravats et al. 2018). We also

find that specific binding partners and environment impacts the direction of the shift of growth rate compared to standard conditions, indicating that different environments place specific functional demands on Hsp90 that could potentially be alleviated by relative affinity of different clients and co-chaperones. In fact, previous studies have identified Hsp90 interactions with specific stress induced clients and co-chaperones that are important for yeast survival and growth when exposed to specific environmental insults, including the stress activated MAPK Hog1 in conjunction with the co-chaperone Cdc37 in *S. cerevisiae* (Hawle et al. 2007), and *C. albicans* (Diezmann et al. 2012), the stress activated mitogen-activated protein kinase, Slt2p in *S. Cerevisiae* (Millson et al. 2005) and calcineurin in *S. cerevisiae* (Imai and Yahara. 2000) Together these results indicate that client and co-chaperone interactions with Hsp90 have a larger impact on experimental fitness under stress in comparison to ATP binding and hydrolysis which are main structural determinants that constrain fitness under standard conditions.

In Chapter II we observed a high correlation of mutant fitness effects across environments indicating that most environments have similar effects on Hsp90 mutations, apart from diamide. Diamide did not show a high correlation of mutant fitness effects across environments suggesting that diamide had distinct effects on the fate of mutations similar to what we observed in Chapter III. The observation that mutant fitness effects correlates across environments contrasts with the distinct fates of mutations in each environment in Chapter III. There are multiple technological and experimental factors that could have implications and limitations that give rise to observed differences. For example, ethanol, high temperature and nitrogen deprivation examined in Chapter III were not included in our analysis in Chapter II, and sorbitol and

H202 examined in Chapter II were not included in Chapter III so we cannot directly compare the overall results of environmental impacts on the DFE of Hsp90 mutants from both studies. Furthermore, in Chapter III we measured the fitness effects of Hsp90 mutations under artificially low expression to tightly couple function of Hsp90 to growth rates and provide a sensitive readout of fitness, whereas, in Chapter II we measured fitness effects of Hsp90 mutations at endogenous expression levels to observe what would happen naturally. Previous studies of low Hsp90 expression identified latent fitness effects and demonstrated that endogenous expression levels obscured fitness effects (Jiang et al. 2013). Therefore, under endogenous Hsp90 expression levels we may be unable to detect similar fitness effects to those observed under low Hsp90 expression. Furthermore, while we found that expression of Hsp90 at low and endogenous expression levels remains stable under conditions examined, this promoter does not capture native transcriptional regulation and thus may not fully recapitulate translational and post-translational regulation controlling Hsp82. As a result, all these factors could have implications on observed results and differences. Future experiments that investigate the DFE of Hsp90 mutations under all conditions and expression levels examined could delineate whether differences between studies are due to these factors or environments examined.

### **The Adaptive Potential of Hsp90 Across Environmental Conditions**

In both studies we detected a small but relative number of beneficial, “adaptive” mutations across all environments except diamide and 37°C, which showed the largest proportion and magnitude of beneficial mutations. The overall results coincides with

observations that beneficial mutations are rare but can provide a selective advantage under experimental conditions (Elena 1998; Thatcher et al. 1998; Sanjuán et al. 2004; Burch et al. 2007; Silander et al. 2007). In Chapter II these beneficial mutations were enriched in a specific region of Hsp90 that coincides with an allosteric center thought to be involved in stabilization of Hsp90 interdomains and stabilization of client-binding interfaces (Blacklock and Verkhivker 2013; Blacklock and Verkhivker 2014).

Furthermore, this beneficial hotspot is adjacent to a known client-binding loop, which suggests that beneficial mutations may promote Hsp90 conformational changes and/or changes in dynamics and stability that may impact nearby client binding affinity and priority of different clients, further facilitating adaptation to specific conditions. Further structural analysis of the biochemical properties of these mutations may provide mechanistic insight into how these mutations impact fitness. The fitness effects of these beneficial mutations in alternate environments identified a region of Hsp90 (amino acids 381-391) with little costs of adaptation. Specifically, we found that on average 14% of beneficial mutations within this region showed deleterious effects in alternate environments, indicating the possibility of little cost of adaptation for mutations within this particular region of Hsp90. This finding was consistent with the high correlation of observed mutant fitness effects across environments but at odds with previous results from a smaller, 9 amino acid client binding loop in the C terminal of Hsp90, which identified high costs of adaptation for beneficial mutations in other environments (Hietpas et al. 2013). Because we examined a relatively larger and different region of Hsp90 this could have multiple implications on the observed minimal costs of adaptation for beneficial mutations. For example, together many positions and beneficial mutations

may have similar roles across environments and may have a large cost of adaptation in other environments, but due to their low proportion in the total number of observed mutations they have a minimum impact on the correlation of selection coefficients across environments. We find that measuring the correlation of a smaller 10 amino acid region across environments shows variability consistent with our explanation as to why we see differences in costs between our and previous studies. To confirm these beneficial mutant fitness effects and their costs in alternate environments, future studies that examine the growth of a panel of individual beneficial mutations in monoculture across environments are warranted.

In Chapter III we found adaptive mutations clustered at specific regions including the ATP domain and middle domain. Further structural analysis revealed that adaptive mutations within these regions were found on the surface of Hsp90 and at positions accessible to binding partners. Because clustered beneficial mutations were preferentially found within these regions, we hypothesize that mutations at these positions may mediate the binding affinity of subsets of clients and co-chaperones and that disruptive mutations can lead to re-prioritization of multiple clients, which may enable priority of efficiency of Hsp90 for specific clients that in turn impact physiology and adaptation to these environments. Further analysis of these beneficial mutations under alternate environments identified high costs of adaptation, consistent with observation of high costs for diamide adaptive mutations in chapter II and previous studies by Hietpas et al., 2013. We also identified strongly beneficial mutations under elevated temperature that are found with the first seven amino acids of Hsp90. Our observations are consistent with previous studies that demonstrate that mRNA structure

adjacent to the start of coding regions impacts translation efficiency (Tuller et al. 2010; Plotkin and Kudla 2011; Li 2015) and the previous observation that changes in expression levels can facilitate adaptation (Lang and Desai 2014).

To gain further insight into the overall adaptive potential derived from the DFE of Hsp90 mutations across environments we interpreted our results considering Fisher's Geometric Model (FGM). Briefly, FGM assumes that populations evolve in a multidimensional geometric phenotypic space through random single step mutations, where the same mutation is more likely to be beneficial for an individual that is far from the phenotypic optimum (Fisher 1931; Tenaillon 2014). As a result, the proportion of beneficial mutations considering FGM expectations should provide information regarding the harshness of the experimental environment and resulting adaptive potential under each environment for the middle domain of Hsp90 (Chapter II). Additionally, the populations distance from the optimum in each environment, using a modified version of FGM (Martin and Lenormand 2006) should provide information regarding the availability of adaptive mutations or adaptive potential of Hsp90 under each environment (Chapter III). Based on FGM we expected to find a larger number of beneficial mutations under environments with large doubling times, corresponding to lower absolute fitness and the lowest number of beneficials under environments that the wild type is well adapted to and has the lowest doubling time, such as under standard conditions. In Chapter II we found that diamide met this expectation, in comparison, high salinity and sorbitol conditions had less beneficial mutations in comparison to standard conditions. Our findings under standard, elevated salinity and sorbitol conditions are at odds with FGM expectations, which could be due to the models

general assumption that mutations affect all phenotypic dimensions with equal weight (Harmand et al. 2017). Because Hsp90 plays a different functional role in osmotic versus oxidative stress this may result in different phenotypic distributions of the same mutations in phenotypic space and thus violates FGM's assumptions. When determining the populations of yeast farthest from the optimum in Chapter III we found that yeast populations under elevated temperature and diamide conditions were farthest from the optimum, followed by nitrogen deprivation, salt, ethanol and standard conditions. These results indicate that diamide and temperature results in the largest potential for adaptation within Hsp90 and is consistent with the largest proportions of beneficial mutations observed in these two environments. The results under elevated temperature were at odds with a previous study of a 9 amino acid client binding loop of Hsp90 by Hietpas et al., 2013, which observed increased availability of adaptive mutations under elevated salinity conditions in comparison to high temperature. These results suggest that the entire Hsp90 protein sequence may present more opportunity for adaptive potential under both temperature and diamide conditions versus a smaller region of Hsp90 that may already be highly optimized for high temperature conditions.

### **Diamide and 37°C Places Distinct Selection Pressures on Hsp90 Mutations**

We observed that diamide in Chapters II and III and heat in Chapter III placed distinct selection pressures on adaptive and deleterious Hsp90 mutations in comparison to other environments. Furthermore, diamide and heat resulted in greater proportions and magnitudes of beneficial and deleterious mutations in comparison to other



environments. The similar results observed under high temperature and diamide in comparison to other environments may be because both stresses may exert similar proteotoxic effects that may disrupt various and or similar cellular processes, ultimately affecting yeast fitness and survival. Furthermore, previous studies by Gasch et al., have identified that both of these stresses result in similar transcriptional profiles in *S. cerevisiae* including the upregulation of Hsp90 gene expression, indicating that these stresses have similar effects on yeast adaptive gene expression response (Gasch et al. 2000). As a result, both these stress conditions may place similar increased selection pressures on Hsp90 function, Hsp90 expression and Hsp90 mutations, in comparison to other environments. For example, because heat results in protein unfolding and protein instability this may result in an increase in Hsp90 unfolded clients that places extra demand on Hsp90 chaperone function and increased selection pressure of Hsp90 mutations that may disrupt Hsp90 function and/or client-binding. This disruption may lead to reprioritization of other clients, which, due to Hsp90s role in multiple signaling pathways has the potential to modify important stress response networks. Consistent with this, we found increased selection of beneficial mutations that corresponds with ATP binding and client-co-chaperone interaction sites in Chapter III. Furthermore, we identified deleterious temperature sensitive (ts) destabilizing mutations within the homodimerization domain of Hsp90, consistent with the requirement for increased folding stability of Hsp90 at high temperatures. We also identified expression dependent ts mutations within surface exposed residues which correlates with reduced Hsp90 function.

The results observed under diamide conditions may be due to the harshness of this oxidant that may result in similar proteotoxic effects. Diamide is a sulfhydryl oxidizing agent that easily penetrates cell membranes and rapidly reacts with low molecular weight thiols to promote intracellular protein disulfide cross-linking (Kosower 1995). A previous study of global changes in protein disulfide bond formation following exposure to diamide identified a number of cytoplasmic disulfide bonded proteins in neuronal cells including, proteins involved in translation, glycolysis, cytoskeletal structure, cell growth and signal transduction kinases (Cumming et al. 2004). Most importantly, they observed that the molecular chaperone proteins (Hsp70, Hsc70 and Hsp90) formed more than one intracellular disulfide bond following exposure to diamide (Cumming et al. 2004). The formation of these disulfide bonds may cause changes in Hsp90 structure, dynamics, stability and interactions with clients, which may have implications on Hsp90 mutant function and chaperone activity similar to heat stress. This concept coincides with previous observations of other oxidizing conditions causing disulfide bond formation in Hsp90 that compromises chaperone activity (Nardai et al. 2000) and the utilization of disulfide bonds to stabilize transient interactions of Hsp90 client-co-chaperone interactions (Southworth and Agard 2011). Furthermore, diamide beneficial mutations in Chapter III were enriched in ATP binding and client interaction sites indicating that diamide impacts chaperone activity and function. In Chapter II, we did not identify specific functional hotspots and found mutations scattered throughout the middle domain. Future studies that identify whether beneficial mutations found in Chapter II coincides with previously identified co-chaperone, client- binding or ATPase activity regions would delineate mechanistically how diamide affects these mutations. In

addition to diamides direct effects on protein stability and function a separate study of the yeast short-term gene expression response to diamide found that diamide resulted in the upregulation of specific subsets of genes including cell wall biosynthesis genes and genes involved in protein secretion and processing in the endoplasmic reticulum, indicating its role in cell wall damage in yeast (Gasch et al. 2000). Furthermore, diamide upregulates the expression of chaperone proteins including, Hsp90, similar to results observed under high temperature (Gasch et al. 2000), suggesting that diamide may result in proteotoxic conditions like heat stress conditions that requires increased demand of Hsp90 function (Gasch et al. 2000). Diamide also affects individual transcription factors differently, for example the Yap1 transcription factor is nuclear localized and active in diamide but cytoplasmic and inactive in H<sub>2</sub>O<sub>2</sub> oxidizing conditions, which may explain why we observed distinct differences in mutant fitness effects under diamide and H<sub>2</sub>O<sub>2</sub> conditions in Chapter II (Gulshan et al. 2011). Together these results suggest that diamide and high temperature effects the folding properties and stability of proteins involved in various cellular processes that may exert multiple effects on the cell. Furthermore, many of the affected proteins and transcription factors are potential Hsp90 clients, which may ultimately place increased selection pressure on Hsp90 function. This corresponds with the observation of increased selection of beneficial mutations found within client-co-chaperone interaction sites in Hsp90 under both diamide and high temperature conditions in our studies. Furthermore our results correlate with previous studies by Gasch et al., that identified increased expression of subsets of genes under both high temperature and diamide conditions including genes that encode for kinases and transcription factors (Gasch et al. 2000). Many of these

upregulated genes were later identified as bona fide Hsp90 clients in studies by McClellan et al., including the transcription factor Yap1 and the kinase Slit2 (McClellan et al. 2007). Because both conditions directly affect Hsp90 expression and function this may result in increased selection of Hsp90 mutations that may prime Hsp90 for these two environmental stresses. Further global experimental analyses beyond the scope of this work will be required to determine the molecular features of diamide and heat stress that elicit distinct selection on Hsp90 mutations.

### **Concluding Remarks and Future Studies**

While there has been studies investigating the effect of environment on the DFE of mutations within genes there are few studies that have looked at how costs of adaptation vary along a large stretch of a protein sequence and to our knowledge there are no experimental studies that have identified the environmental dependence of most mutations in an entire protein sequence for a large protein. The work presented here is a comprehensive interrogation of the effects of environment on the DFE of mutations within a large client- binding region of the Hsp90 sequence under endogenous expression levels and the entire Hsp90 sequence under low expression levels. Our work is the first to present comprehensive fitness maps for an entire protein sequence and a large functional region of a protein sequence across environments. This thesis elucidates the impact environment has on shaping experimental and natural Hsp90 evolution and uncovers adaptive potential, costs of adaptation and evolutionary constraints for Hsp90 mutations across environments. My experimental studies have shed light on how specific environments can impact the selection of Hsp90 mutations

and their fate in other environments, which can ultimately shape Hsp90 evolution. Further comparison between experimental fitness effects of mutations and natural Hsp90 variants across environments highlights that fluctuating environments may have contributed to the long-term evolution of natural Hsp90 variants. In addition, my studies have shed light on potential molecular mechanisms of adaptation for beneficial mutations including, potential changes in Hsp90 co-chaperone and client-binding interactions, ATPase activity, structural dynamics and stability.

Future studies for following up on this work would be to determine the molecular features of diamide and heat stress that elicit distinct selection of beneficial Hsp90 mutations in comparison to standard conditions. Because of the numerous and transient nature of Hsp90 potential interactions with co-chaperones and clients, standard analytical techniques such as, molecular dynamic simulations, crystallography, NMR, and proteomic studies are not ideal. Indirect approaches that include a global analysis of beneficial Hsp90 mutants and wild-type Hsp90 mRNA levels under diamide, heat stress and standard conditions combined with bioinformatics may provide an opportunity to discover potential Hsp90 mutant-client interactions, pathways and cellular functions that may be affected in each condition. Specifically, using mRNA Seq combined with Gene Ontology Enrichment and network analysis I can systematically compare the expression levels of each gene, determine what genes are upregulated or downregulated, identify whether these genes are potential Hsp90 clients and ultimately identify the pathways and cellular functions that may be affected in each condition. For example, previous studies by McClellan et al., used genome-wide chemical-genetic screens in *S. cerevisiae* grown under standard and heat stress conditions, combined

with GO Enrichment and network analysis (McClellan et al. 2007) to characterize novel Hsp90 clients and novel Hsp90 cellular functions in each environment (McClellan et al. 2007). From this initial screen I can pick potential Hsp90 clients for further systematic analysis of their binding interactions with wild-type Hsp90 and Hsp90 mutants across environments using the LUMIER binding interaction assay (Taipale et al. 2012). The LUMIER binding interaction assay has been previously used to systematically identify all human Hsp90 kinase interactions (Taipale et al. 2012). This systematic analysis should ultimately identify Hsp90 mutant-client binding interactions among environments and highlight whether these adaptive Hsp90 mutants interact with similar, novel and or different clients. Finally, mechanistic studies that include: client maturation assays to investigate a panel of adaptive Hsp90 mutants and their effects on the maturation of model yeast clients including GR, v-SRC and Ste11 could identify mutations that impact client maturation; ATPase activity assays of adaptive Hsp90 mutants could determine the ATPase activity of these mutants and identify mutations that impact Hsp90 ATPase chaperone function; and single molecule analysis, such as FRET of Hsp90 adaptive mutants could help identify mutations that impact Hsp90 stability and or dynamics. Together these studies should help delineate how each adaptive mutation impacts Hsp90 and provide further mechanistic insight into the adaptive potential of Hsp90.

## References

- Abecasis, G. R., A. Auton, L. D. Brooks, M. A. DePristo, R. M. Durbin, R. E. Handsaker, H. M. Kang, G. T. Marth, G. A. McVean. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491:56-65.
- Abriata, L. A., T. Palzkill, and M. Dal Peraro. 2015. How structural and physicochemical determinants shape sequence constraints in a functional enzyme. *PLoS One* 10:e0118684-e0118684.
- Acevedo, A., L. Brodsky, and R. Andino. 2014. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* 505:686-690.
- Aharoni, A., L. Gaidukov, O. Khersonsky, Q. G. S. Mc, C. Roodveldt, and D. S. Tawfik. 2005. The 'evolvability' of promiscuous protein functions. *Nat. Genet.* 37:73-76.
- Aita, T., N. Hamamatsu, Y. Nomiya, H. Uchiyama, Y. Shibanaka, and Y. Husimi. 2002. Surveying a local fitness landscape of a protein with epistatic sites for the study of directed evolution. *Biopolymers* 64:95-105.
- Ali, M. M. U., S. M. Roe, C. K. Vaughan, P. Meyer, B. Panaretou, P. W. Piper, C. Prodromou, and L. H. Pearl. 2006. Crystal structure of an Hsp90–nucleotide–p23/Sba1 closed chaperone complex. *Nature* 440:1013-1017.
- Anderson, S. 1981. Shotgun DNA sequencing using cloned DNase I-generated fragments. *Nucleic Acids Res.* 9:3015-3027.
- Araya, C. L., D. M. Fowler, W. Chen, I. Muniez, J. W. Kelly, and S. Fields. 2012. A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. *Proceedings of the National Academy of Sciences* 109:16858-16863.
- Arribas, M., K. Kubota, L. Cabanillas, and E. Lazaro. 2014. Adaptation to fluctuating temperatures in an RNA virus is driven by the most stringent selective pressure. *PLoS One* 9:e100940.
- Ashkenazy, H., S. Abadi, E. Martz, O. Chay, I. Mayrose, T. Pupko, and N. Ben-Tal. 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* 44:W344-W350.
- Ayala, F. J. and W. M. Fitch. 1997. Genetics and the origin of species: An introduction. *Proceedings of the National Academy of Sciences* 94:7691-7697.
- Bailey, S. F. and T. Bataillon. 2016. Can the experimental evolution programme help us elucidate the genetic basis of adaptation in nature? *Mol. Ecol.* 25:203-218.
- Bank, C., R. T. Hietpas, J. D. Jensen, and D. N. A. Bolon. 2015. A systematic survey of an intragenic epistatic landscape. *Mol. Biol. Evol.* 32:229-238.
- Bank, C., R. T. Hietpas, A. Wong, D. N. Bolon, and J. D. Jensen. 2014. A Bayesian MCMC Approach to Assess the Complete Distribution of Fitness Effects of New Mutations: Uncovering the Potential for Adaptive Walks in Challenging Environments. *Genetics* 196:841-852.
- Barrett, R. D. and D. Schluter. 2008. Adaptation from standing genetic variation. *Trends Ecol. Evol.* 23:38-44.
- Barton, J. P., M. Kardar, and A. K. Chakraborty. 2015. Scaling laws describe memories of host–pathogen riposte in the HIV population. *Proceedings of the National Academy of Sciences* 112:1965-1970.
- Bataillon, T. and S. F. Bailey. 2014. Effects of new mutations on fitness: insights from models and data. *Ann. N. Y. Acad. Sci.* 1320:76-92.
- Bataillon, T., T. Zhang, and R. Kassen. 2011. Cost of adaptation and fitness effects of beneficial mutations in *Pseudomonas fluorescens*. *Genetics* 189:939-949.
- Bateson, W. 1894. Materials for the study of variation, treated with especial regard to discontinuity in the origin of species 1861–1926. London : Macmillan

- Bershtein, S., W. Mu, A. W. R. Serohijos, J. Zhou, and E. I. Shakhnovich. 2013. Protein quality control acts on folding intermediates to shape the effects of mutations on organismal fitness. *Mol. Cell* 49:133-144.
- Biesecker, B. B. and H. L. Peay. 2013. Genomic sequencing for psychiatric disorders: promise and challenge. *The international journal of neuropsychopharmacology* 16:1667-1672.
- Biesecker, L. G. 2010. Exome sequencing makes medical genomics a reality. *Nat. Genet.* 42:13-14.
- Blacklock, K. and G. Verkhivker. 2013. Differential Modulation Of Functional Dynamics and Allosteric Interactions In the Hsp90-Cochaperone Complexes With P23 and Aha1: A Computational Study. *PLoS One* 8:e71936.
- Blacklock, K. and G. M. Verkhivker. 2014. Allosteric regulation of the Hsp90 dynamics and stability by client recruiter cochaperones: protein structure network modeling. *PLoS One* 9:e86547-e86547.
- Bloom, J. D. and F. H. Arnold. 2009. In the light of directed evolution: Pathways of adaptive protein evolution. *Proceedings of the National Academy of Sciences* 106:9995-10000.
- Blount, Z. D., J. E. Barrick, C. J. Davidson, and R. E. Lenski. 2012. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* 489:513-518.
- Blount, Z. D., C. Z. Borland, and R. E. Lenski. 2008. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proceedings of the National Academy of Sciences* 105:7899-7906.
- Boder, E. T. and K. D. Wittrup. 1997. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* 15:553-557.
- Bohen, S. P. and K. R. Yamamoto. 1993. Isolation of Hsp90 mutants by screening for decreased steroid receptor function. *Proc. Natl. Acad. Sci. U. S. A.* 90:11424-11428.
- Borkovich, K. A., F. W. Farrelly, D. B. Finkelstein, J. Taulien, and S. Lindquist. 1989. hsp82 is an essential protein that is required in higher concentrations for growth of cells at higher temperatures. *Mol. Cell. Biol.* 9:3919-3930.
- Botstein, D. and D. Shortle. 1985. Strategies and applications of in vitro mutagenesis. *Science* 229:1193-1201.
- Boucher, J. I., D. N. Bolon, and D. S. Tawfik. 2016. Quantifying and understanding the fitness effects of protein mutations: Laboratory versus nature. *Protein Sci.* 25:1219-1226.
- Boucher, J. I., P. Cote, J. Flynn, L. Jiang, A. Laban, P. Mishra, B. P. Roscoe, and D. N. Bolon. 2014. Viewing protein fitness landscapes through a next-gen lens. *Genetics* 198:461-471.
- Boucher, J. I., T. W. Whitfield, A. Dauphin, G. Nachum, C. Hollins, K. B. Zeldovich, R. Swanstrom, C. A. Schiffer, J. Luban, and D. N. A. Bolon. 2019. Constrained Mutational Sampling of Amino Acids in HIV-1 Protease Evolution. *Mol. Biol. Evol.* 36:798-810.
- Boyko, A. R., S. H. Williamson, A. R. Indap, J. D. Degenhardt, R. D. Hernandez, K. E. Lohmueller, M. D. Adams, S. Schmidt, J. J. Sninsky, S. R. Sunyaev, T. J. White, R. Nielsen, A. G. Clark, and C. D. Bustamante. 2008. Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet.* 4:e1000083-e1000083.
- Brennan, G. L., N. Colegrave, and S. Collins. 2017. Evolutionary consequences of multidriver environmental change in an aquatic primary producer. *Proc. Natl. Acad. Sci. U. S. A.* 114:9930-9935.
- Brockhurst, M. A., N. Colegrave, and D. E. Rozen. 2011. Next-generation sequencing as a tool to study microbial evolution. *Mol. Ecol.* 20:972-980.
- Bui, D. C., Y. Lee, J. Y. Lim, M. Fu, J. C. Kim, G. J. Choi, H. Son, and Y. W. Lee. 2016. Heat shock protein 90 is required for sexual and asexual development, virulence, and heat shock response in *Fusarium graminearum*. *Sci. Rep.* 6:28154.
- Burch, C. L., S. Guyader, D. Samarov, and H. Shen. 2007. Experimental estimate of the abundance and effects of nearly neutral mutations in the RNA virus phi 6. *Genetics* 176:467-476.



- Cadwell RC, J. G. 1994. Mutagenic PCR. *PCR Methods Appl.* 3.
- Canale, A. S., S. V. Venev, T. W. Whitfield, D. R. Caffrey, W. A. Marasco, C. A. Schiffer, T. F. Kowalik, J. D. Jensen, R. W. Finberg, K. B. Zeldovich, J. P. Wang, and D. N. A. Bolon. 2018. Synonymous Mutations at the Beginning of the Influenza A Virus Hemagglutinin Gene Impact Experimental Fitness. *J. Mol. Biol.* 430:1098-1115.
- Chakravarty S, V. R. 1999. Residue depth: a novel parameter for the analysis of protein structure and stability. *Structure* 7:.
- Chang, H. C. and S. Lindquist. 1994. Conservation of Hsp90 macromolecular complexes in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 269:24983-24988.
- Charlesworth, D., N. H. Barton, and B. Charlesworth. 2017. The sources of adaptive variation. *Proc. Biol. Sci.* 284.
- Charlesworth, D., B. Charlesworth, and M. T. Morgan. 1995. The pattern of neutral molecular variation under the background selection model. *Genetics* 141:1619-1632.
- Chen, K. and F. H. Arnold. 1993. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc. Natl. Acad. Sci. U. S. A.* 90:5618-5622.
- Chockalingam, K., Z. Chen, J. A. Katzenellenbogen, and H. Zhao. 2005. Directed evolution of specific receptor-ligand pairs for use in the creation of gene switches. *Proc. Natl. Acad. Sci. U. S. A.* 102:5691-5696.
- Cobb, M. 2017. 60 years ago, Francis Crick changed the logic of biology. *PLoS Biol.* 15:e2003243.
- Cobb, R. E., R. Chao, and H. Zhao. 2013. Directed Evolution: Past, Present and Future. *AIChE journal. American Institute of Chemical Engineers* 59:1432-1440.
- Cooper, G. M., D. L. Goode, S. B. Ng, A. Sidow, M. J. Bamshad, J. Shendure, and D. A. Nickerson. 2010. Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. *Nature methods* 7:250-251.
- Corcos, A. F. and F. V. Monaghan. 1987. Correns, an independent discoverer of Mendelism?: I. An historical/critical note. *J. Hered.* 78:330-330.
- Correns, C. G. 1950. Mendel's Law Concerning the Behavior of Progeny of Varietal Hybrids. *Genetics* 35.
- Cowen, L. E. and S. Lindquist. 2005. Hsp90 potentiates the rapid evolution of new traits: drug resistance in diverse fungi. *Science* 309:2185-2189.
- Crick, F. 1970. Central Dogma of Molecular Biology. *Nature* 227:561-563.
- Crick, F. H. 1958. On protein synthesis. *Symp. Soc. Exp. Biol.* 12:138-163.
- Cumming, R., N. Andon, P. Haynes, M. Park, W. Fischer, and D. Schubert. 2004. Protein Disulfide Bond Formation in the Cytoplasm during Oxidative Stress. *The Journal of biological chemistry* 279:21749-21758.
- Czemerer, J., K. Buse, and G. M. Verkhivker. 2017. Atomistic simulations and network-based modeling of the Hsp90-Cdc37 chaperone binding with Cdk4 client protein: A mechanism of chaperoning kinase clients by exploiting weak spots of intrinsically dynamic kinase domains. *PLoS One* 12:e0190267.
- Dandage, R., R. Pandey, G. Jayaraj, and K. Chakraborty. 2017. Differential strengths of molecular determinants guide environment specific mutational fates. *bioRxiv:134569*.
- Darwin, C. 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life.* (Down, Bromley, Kent), .
- Darwin, C. and A. Wallace. 1858. On the Tendency of Species to form Varieties; and on the Perpetuation of Varieties and Species by Natural Means of Selection. *Journal of the Proceedings of the Linnean Society of London. Zoology* 3:45-62.
- de Vries, H. 1950. Concerning the Law of Segregation of Hybrids. *Genetics* 35.

- Dehouck, Y., J. M. Kwasigroch, D. Gilis, and M. Rومان. 2011. PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC Bioinformatics* 12:151-151.
- Dhar, R., R. Sagesser, C. Weikert, J. Yuan, and A. Wagner. 2011. Adaptation of *Saccharomyces cerevisiae* to saline stress through laboratory evolution. *J. Evol. Biol.* 24:1135-1153.
- Diezmann, S., M. Michaut, R. S. Shapiro, G. D. Bader, and L. E. Cowen. 2012. Mapping the Hsp90 genetic interaction network in *Candida albicans* reveals environmental contingency and rewired circuitry. *PLoS Genet.* 8:e1002562.
- Dillon, M. M., N. P. Rouillard, B. Van Dam, R. Gallet, and V. S. Cooper. 2016. Diverse phenotypic and genetic responses to short-term selection in evolving *Escherichia coli* populations. *Evolution* 70:586-599.
- Dingens, A. S., D. Arenz, J. Overbaugh, and J. D. Bloom. 2018. Massively parallel profiling of HIV-1 resistance to the fusion inhibitor enfuvirtide. *bioRxiv:472746*.
- Dobzhansky, T. 1937. *Genetics and the Origin of Species*. (Columbia Univ. Press, New York); 2nd Ed., 1941; 3rd Ed., 1951.
- Doud, M. B. and J. D. Bloom. 2016. Accurate Measurement of the Effects of All Amino-Acid Mutations on Influenza Hemagglutinin. *Viruses* 8:155.
- Duch, A., E. de Nadal, and F. Posas. 2012. The p38 and Hog1 SAPKs control cell cycle progression in response to environmental stresses. *FEBS Lett.* 586:2925-2931.
- Duenas-Decamp, M., L. Jiang, D. Bolon, and P. R. Clapham. 2016. Saturation Mutagenesis of the HIV-1 Envelope CD4 Binding Loop Reveals Residues Controlling Distinct Trimer Conformations. *PLoS Pathog.* 12:e1005988.
- Dykhuizen, D. E., A. M. Dean, and D. L. Hartl. 1987. Metabolic flux and fitness. *Genetics* 115:25-31.
- Elena, S. F., Ekunwe, L., Hajela, N., Oden, S. A. & Lenski, R. E. . 1998. Distribution of fitness effects caused by random insertion mutations in *Escherichia coli*. *Genetica* 102–103,.
- Elena, S. F. and R. E. Lenski. 2003. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature reviews. Genetics* 4:457-469.
- Eyre-Walker, A. and P. D. Keightley. 2007. The distribution of fitness effects of new mutations. *Nature Reviews Genetics* 8:610-618.
- Ferea, T. L., D. Botstein, P. O. Brown, and R. F. Rosenzweig. 1999. Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 96:9721-9726.
- Firnberg, E., J. W. Labonte, J. J. Gray, and M. Ostermeier. 2014. A comprehensive, high-resolution map of a gene's fitness landscape. *Mol. Biol. Evol.* 31:1581-1592.
- Fisher, R. A., . . 1958. *The genetical theory of natural selection*.
- Fisher, R. A. S. 1931. *The genetical theory of natural selection*. Oxford University Press, Oxford.
- Fitzgerald, D. M. and S. M. Rosenberg. 2019. What is mutation? A chapter in the series: How microbes "jeopardize" the modern synthesis. *PLoS Genet.* 15:e1007995.
- Flynn, J. M., A. Rossouw, P. Cote-Hammarlof, I. Fragata, D. Mavor, C. Hollins, 3rd, C. Bank, and D. N. Bolon. 2020. Comprehensive fitness maps of Hsp90 show widespread environmental dependence. *Elife* 9.
- Fowler, D. M., C. L. Araya, S. J. Fleishman, E. H. Kellogg, J. J. Stephany, D. Baker, and S. Fields. 2010. High-resolution mapping of protein sequence-function relationships. *Nature methods* 7:741-746.
- Fowler, D. M. and S. Fields. 2014. Deep mutational scanning: a new style of protein science. *Nature methods* 11:801-807.
- Fragata, I., S. Matuszewski, M. A. Schmitz, T. Bataillon, J. D. Jensen, and C. Bank. 2018. The fitness landscape of the codon space across environments. *Heredity (Edinb.)* 121:422-437.
- Freese, E. B., M. I. Chu, and E. Freese. 1982. Initiation of yeast sporulation of partial carbon, nitrogen, or phosphate deprivation. *J. Bacteriol.* 149:840-851.

- Gasch, A. P., P. T. Spellman, C. M. Kao, O. Carmel-Harel, M. B. Eisen, G. Storz, D. Botstein, and P. O. Brown. 2000. Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* 11:4241-4257.
- Genest, O., M. Reidy, T. O. Street, J. R. Hoskins, J. L. Camberg, D. A. Agard, D. C. Masison, and S. Wickner. 2013. Uncovering a region of heat shock protein 90 important for client binding in *E. coli* and chaperone function in yeast. *Mol. Cell* 49:464-473.
- Gietz, R. D. and R. H. Schiestl. 2007. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* 2:31-34.
- Gillham, N. W. 2001. Evolution by Jumps: Francis Galton and William Bateson and the Mechanism of Evolutionary Change. *Genetics* 159:1383.
- Girstmair, H., F. Toppel, A. Lopez, K. Tych, F. Stein, P. Haberkant, P. W. N. Schmid, D. Helm, M. Rief, M. Sattler, and J. Buchner. 2019. The Hsp90 isoforms from *S. cerevisiae* differ in structure, function and client range. *Nature communications* 10:3626.
- González-Pérez, A. and N. López-Bigas. 2011. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am. J. Hum. Genet.* 88:440-449.
- Gorter, F. A., M. F. L. Derks, J. van den Heuvel, M. G. M. Aarts, B. J. Zwaan, D. de Ridder, and J. de Visser. 2017. Genomics of Adaptation Depends on the Rate of Environmental Change in Experimental Yeast Populations. *Mol. Biol. Evol.* 34:2613-2626.
- Gulshan, K., S. S. Lee, and W. S. Moye-Rowley. 2011. Differential oxidant tolerance determined by the key transcription factor Yap1 is controlled by levels of the Yap1-binding protein, Ybp1. *J. Biol. Chem.* 286:34071-34081.
- Haber, J. E. 2012. Mating-type genes and MAT switching in *Saccharomyces cerevisiae*. *Genetics* 191:33-64.
- Hagn, F., S. Lagleder, M. Retzlaff, J. Rohrberg, O. Demmer, K. Richter, J. Buchner, and H. Kessler. 2011. Structural analysis of the interaction between Hsp90 and the tumor suppressor protein p53. *Nat. Struct. Mol. Biol.* 18:1086-1093.
- Haldane, J. B. S. 1927. A Mathematical Theory of Natural and Artificial Selection, Part V: Selection and Mutation. *Mathematical Proceedings of the Cambridge Philosophical Society* 23:838-844.
- Haldane, J. B. S. 1959. The Theory of Natural Selection To-Day. *Nature* 183:710-713.
- Harmand, N., R. Gallet, R. Jabbour-Zahab, G. Martin, and T. Lenormand. 2017. Fisher's geometrical model and the mutational patterns of antibiotic resistance across dose gradients. *Evolution* 71:23-37.
- Hawle, P., D. Horst, J. P. Bebelman, X. X. Yang, M. Siderius, and S. M. van der Vies. 2007. Cdc37p is required for stress-induced high-osmolarity glycerol and protein kinase C mitogen-activated protein kinase pathway functionality by interaction with Hog1p and Slt2p (Mpk1p). *Eukaryotic cell* 6:521-532.
- Hawle, P., M. Siepmann, A. Harst, M. Siderius, H. P. Reusch, and W. M. Obermann. 2006. The middle domain of Hsp90 acts as a discriminator between different types of client proteins. *Mol. Cell. Biol.* 26:8385-8395.
- Heather, J. M. and B. Chain. 2016. The sequence of sequencers: The history of sequencing DNA. *Genomics* 107:1-8.
- Heberle, H., G. V. Meirelles, F. R. da Silva, G. P. Telles, and R. Minghim. 2015. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* 16:169.
- Henikoff, S. and J. G. Henikoff. 1992. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U. S. A.* 89:10915-10919.
- Hiatt, J. B., R. P. Patwardhan, E. H. Turner, C. Lee, and J. Shendure. 2010. Parallel, tag-directed assembly of locally derived short sequence reads. *Nature methods* 7:119-122.

- Hietpas, R., B. Roscoe, L. Jiang, and D. N. Bolon. 2012. Fitness analyses of all possible point mutations for regions of genes in yeast. *Nat. Protoc.* 7:1382-1396.
- Hietpas, R. T., C. Bank, J. D. Jensen, and D. N. A. Bolon. 2013. Shifting fitness landscapes in response to altered environments. *Evolution* 67:3512-3522.
- Hietpas, R. T., J. D. Jensen, and D. N. Bolon. 2011. Experimental illumination of a fitness landscape. *Proc. Natl. Acad. Sci. U. S. A.* 108:7896-7901.
- Hill, J. A. and S. P. Otto. 2007. The role of pleiotropy in the maintenance of sex in yeast. *Genetics* 175:1419-1427.
- Hirata, Y., T. Andoh, T. Asahara, and A. Kikuchi. 2003. Yeast glycogen synthase kinase-3 activates Msn2p-dependent transcription of stress responsive genes. *Mol. Biol. Cell* 14:302-312.
- Holley, R. W., J. Apgar, G. A. Everett, J. T. Madison, M. Marquisee, S. H. Merrill, J. R. Penswick, and A. Zamir. 1965. Structure of a Ribonucleic Acid. *Science* 147:1462-1465.
- Hom, N., L. Gentles, J. D. Bloom, and K. K. Lee. 2019. Deep Mutational Scan of the Highly Conserved Influenza A Virus M1 Matrix Protein Reveals Substantial Intrinsic Mutational Tolerance. *J. Virol.* 93.
- Huxley, J. 1944. *Evolution: The Modern Synthesis*. Julian Huxley. *Isis* 35:192-194.
- Imai, J. and I. Yahara. 2000. Role of HSP90 in salt stress tolerance via stabilization and regulation of calcineurin. *Mol. Cell. Biol.* 20:9262-9270.
- Jarosz, D. F. and S. Lindquist. 2010. Hsp90 and environmental stress transform the adaptive value of natural genetic variation. *Science* 330:1820-1824.
- Jiang, L., P. Liu, C. Bank, N. Renzette, K. Prachanonrongs, L. S. Yilmaz, D. R. Caffrey, K. B. Zeldovich, C. A. Schiffer, T. F. Kowalik, J. D. Jensen, R. W. Finberg, J. P. Wang, and D. N. A. Bolon. 2016. A Balance between Inhibitor Binding and Substrate Processing Confers Influenza Drug Resistance. *J. Mol. Biol.* 428:538-553.
- Jiang, L., P. Mishra, R. T. Hietpas, K. B. Zeldovich, and D. N. A. Bolon. 2013. Latent effects of Hsp90 mutants revealed at reduced expression levels. *PLoS Genet.* 9:e1003600-e1003600.
- Kacser, H. and J. A. Burns. 1981. The molecular basis of dominance. *Genetics* 97:639-666.
- Kaplan, K. B. and R. Li. 2012. A prescription for 'stress'--the role of Hsp90 in genome stability and cellular adaptation. *Trends Cell Biol.* 22:576-583.
- Karagoz, G. E., A. M. Duarte, E. Akoury, H. Ippel, J. Biernat, T. Moran Luengo, M. Radli, T. Didenko, B. A. Nordhues, D. B. Veprintsev, C. A. Dickey, E. Mandelkow, M. Zweckstetter, R. Boelens, T. Madl, and S. G. Rudiger. 2014. Hsp90-Tau complex reveals molecular basis for specificity in chaperone action. *Cell* 156:963-974.
- Kassen, R. and T. Bataillon. 2006. Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. *Nat. Genet.* 38:484-488.
- Kaufman, S. and S. Levin. 1987. Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.* 128:11-45.
- Keightley, P. D. and A. Eyre-Walker. 2010. What can we learn about the distribution of fitness effects of new mutations from DNA sequence data? *Philosophical Transactions of the Royal Society B: Biological Sciences* 365:1187-1193.
- Kemble, H., P. Nghe, and O. Tenailon. 2019. Recent insights into the genotype-phenotype relationship from massively parallel genetic assays. *Evolutionary Applications* 12:1721-1742.
- Khan, A. I., D. M. Dinh, D. Schneider, R. E. Lenski, and T. F. Cooper. 2011. Negative Epistasis Between Beneficial Mutations in an Evolving Bacterial Population. *Science* 332:1193-1196.
- Khersonsky, O. and D. S. Tawfik. 2010. Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* 79:471-505.
- Kim, I., C. R. Miller, D. L. Young, and S. Fields. 2013. High-throughput analysis of in vivo protein stability. *Mol. Cell. Proteomics* 12:3370-3378.

- Kimura, J. F. C., Motoo. . 1970. An introduction to population genetics theory ([Reprint] ed.). New Jersey: Blackburn Press. p. 5. ISBN 978-1-932846-12-6.
- Kimura, M. 1968. Evolutionary Rate at the Molecular Level. *Nature* 217:624.
- Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge University Press.
- Kircher, M., D. M. Witten, P. Jain, B. J. O'Roak, G. M. Cooper, and J. Shendure. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* 46:310-315.
- Kondrashov, D. A. and F. A. Kondrashov. 2015. Topological features of rugged fitness landscapes in sequence space. *Trends Genet.* 31:24-33.
- Kosower, N. S., and Kosower, E. M. 1995. *Methods Enzymol.* 251, 123-13.
- Kouyos, R. D., G. E. Leventhal, T. Hinkley, M. Haddad, J. M. Whitcomb, C. J. Petropoulos, and S. Bonhoeffer. 2012. Exploring the Complexity of the HIV-1 Fitness Landscape. *PLoS Genet.* 8:e1002551.
- Kravats, A. N., J. R. Hoskins, M. Reidy, J. L. Johnson, S. M. Doyle, O. Genest, D. C. Masison, and S. Wickner. 2018. Functional and physical interaction between yeast Hsp90 and Hsp70. *Proc. Natl. Acad. Sci. U. S. A.* 115:E2210-E2219.
- Kresge, N., R. D. Simoni, and R. L. Hill. 2009. Total Synthesis of a Tyrosine Suppressor tRNA: the Work of H. Gobind Khorana. *The Journal of Biological Chemistry* 284:e5-e5.
- Krukenberg, K. A., T. O. Street, L. A. Lavery, and D. A. Agard. 2011. Conformational dynamics of the molecular chaperone Hsp90. *Q. Rev. Biophys.* 44:229-255.
- Kryazhimskiy, S., D. P. Rice, E. R. Jerison, and M. M. Desai. 2014. Microbial evolution. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science (New York, N.Y.)* 344:1519-1522.
- Kvitek, D. J. and G. Sherlock. 2013. Whole genome, whole population sequencing reveals that loss of signaling networks is the major adaptive strategy in a constant environment. *PLoS Genet.* 9:e1003972.
- Laehnemann, D., A. Borkhardt, and A. C. McHardy. 2016. Denoising DNA deep sequencing data-high-throughput sequencing errors and their correction. *Brief. Bioinform.* 17:154-179.
- Lamarck, J. B. 1914. *Zoological Philosophy*. London
- Lamoth, F., P. R. Juvvadi, J. R. Fortwendel, and W. J. Steinbach. 2012. Heat shock protein 90 is required for conidiation and cell wall integrity in *Aspergillus fumigatus*. *Eukaryotic cell* 11:1324-1332.
- Lang, G. I. and M. M. Desai. 2014. The spectrum of adaptive mutations in experimental evolution. *Genomics* 104:412-416.
- Lang, G. I. and A. W. Murray. 2008. Estimating the per-base-pair mutation rate in the yeast *Saccharomyces cerevisiae*. *Genetics* 178:67-82.
- Lang, G. I., D. P. Rice, M. J. Hickman, E. Sodergren, G. M. Weinstock, D. Botstein, and M. M. Desai. 2013. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* 500:571-574.
- Leach, M. D., E. Klipp, L. E. Cowen, and A. J. P. Brown. 2012. Fungal Hsp90: a biological transistor that tunes cellular outputs to thermal inputs. *Nature Reviews Microbiology* 10:693-704.
- Lee, B. and F. M. Richards. 1971. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* 55:379-400.
- Lenski, R. E. 2017. What is adaptation by natural selection? Perspectives of an experimental microbiologist. *PLoS Genet.* 13:e1006668.
- Lenski, R. E. 2020. "We Interrupt This Experiment". *Telliamed Revisited*. Retrieved 2020-03-09. .
- Lenski, R. E., M. R. Rose, S. C. Simpson, and S. C. Tadler. 1991. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *The American Naturalist* 138:1315-1341.

- Lenski, R. E. and M. Travisano. 1994. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proc. Natl. Acad. Sci. U. S. A.* 91:6808-6814.
- Lenski, R. E., M. J. Wiser, N. Ribeck, Z. D. Blount, J. R. Nahum, J. J. Morris, L. Zaman, C. B. Turner, B. D. Wade, R. Maddamsetti, A. R. Burmeister, E. J. Baird, J. Bundy, N. A. Grant, K. J. Card, M. Rowles, K. Weatherspoon, S. E. Papoulis, R. Sullivan, C. Clark, J. S. Mulka, and N. Hajela. 2015. Sustained fitness gains and variability in fitness trajectories in the long-term evolution experiment with *Escherichia coli*. *Proceedings. Biological sciences* 282:20152292-20152292.
- Lerner, S. A., T. T. Wu, and E. C. Lin. 1964. EVOLUTION OF A CATABOLIC PATHWAY IN BACTERIA. *Science* 146:1313-1315.
- Levin, A. M. and G. A. Weiss. 2006. Optimizing the affinity and specificity of proteins with molecular display. *Mol. Biosyst.* 2:49-57.
- Lewontin, R. C. and J. L. Hubby. 1966. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* 54:595-609.
- Li, C. and J. Zhang. 2018. Multi-environment fitness landscapes of a tRNA gene. *Nat Ecol Evol* 2:1025-1032.
- Li, G.-W. 2015. How do bacteria tune translation efficiency? *Curr. Opin. Microbiol.* 24:66-71.
- Li, J., Soroka, J. & Buchner, J. (2012). The Hsp90 chaperone machinery: conformational dynamics and regulation by co-chaperones. *Biochim Biophys Acta* 1823, 624-35. 2012. The Hsp90 chaperone machinery: conformational dynamics and regulation by co-chaperones. *Biochim Biophys Acta* 1823, .
- Lindquist, S. 1981. Regulation of protein synthesis during heat shock. *Nature* 293:311-314.
- Lindquist, S. 2009. Protein folding sculpting evolutionary change. *Cold Spring Harb. Symp. Quant. Biol.* 74:103-108.
- Liu, L., Y. Li, S. Li, N. Hu, Y. He, R. Pong, D. Lin, L. Lu, and M. Law. 2012. Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* 2012:251364-251364.
- Loewe, L. and W. G. Hill. 2010. The population genetics of mutations: good, bad and indifferent. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 365:1153-1167.
- Logacheva, M. D., M. I. Schelkunov, V. Y. Shtratnikova, M. V. Matveeva, and A. A. Penin. 2016. Comparative analysis of plastid genomes of non-photosynthetic Ericaceae and their photosynthetic relatives. *Sci. Rep.* 6:30042.
- Lorenz, O., L. Freiburger, D. Rutz, M. Krause, B. Zierer, S. Alvira de Celis, J. Cuéllar, J. Valpuesta, T. Madl, M. Sattler, and J. Buchner. 2014. Modulation of the Hsp90 Chaperone Cycle by a Stringent Client Protein. *Mol. Cell* 53.
- Lotz, G. P., H. Lin, A. Harst, and W. M. Obermann. 2003. Aha1 binds to the middle domain of Hsp90, contributes to client protein activation, and stimulates the ATPase activity of the molecular chaperone. *J. Biol. Chem.* 278:17228-17235.
- Louvion, J. F., R. Warth, and D. Picard. 1996. Two eukaryote-specific regions of Hsp82 are dispensable for its viability and signal transduction functions in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 93:13937-13942.
- Luria, S. E. and M. Delbrück. 1943. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics* 28:491-511.
- Lynch, M. 2013. Evolutionary diversification of the multimeric states of proteins. *Proc. Natl. Acad. Sci. U. S. A.* 110:E2821-E2828.
- Mager, W. H. and M. Siderius. 2002. Novel insights into the osmotic stress response of yeast. *FEMS yeast research* 2:251-257.

- Martin, G. and T. Lenormand. 2006. A General Multivariate Extension of Fisher's Geometrical Model and the Distribution of Mutation Fitness Effects Across Species. *Evolution* 60:893-907.
- Mattheakis, L. C., R. R. Bhatt, and W. J. Dower. 1994. An in vitro polysome display system for identifying ligands from very large peptide libraries. *Proc. Natl. Acad. Sci. U. S. A.* 91:9022-9026.
- Matuszewski, S., M. E. Hildebrandt, A.-H. Ghenu, J. D. Jensen, and C. Bank. 2016. A Statistical Guide to the Design of Deep Mutational Scanning Experiments. *Genetics* 204:77-87.
- Mavor, D., K. Barlow, S. Thompson, B. A. Barad, A. R. Bonny, C. L. Cario, G. Gaskins, Z. Liu, L. Deming, S. D. Axen, E. Caceres, W. Chen, A. Cuesta, R. E. Gate, E. M. Green, K. R. Hulce, W. Ji, L. R. Kenner, B. Mensa, L. S. Morinishi, S. M. Moss, M. Mravic, R. K. Muir, S. Niekamp, C. I. Nnadi, E. Palovcak, E. M. Poss, T. D. Ross, E. C. Salcedo, S. K. See, M. Subramaniam, A. W. Wong, J. Li, K. S. Thorn, S. Ó. Conchúir, B. P. Roscoe, E. D. Chow, J. L. DeRisi, T. Kortemme, D. N. Bolon, and J. S. Fraser. 2016. Determination of ubiquitin fitness landscapes under different chemical stresses in a classroom setting. *Elife* 5:e15802.
- McCarthy, M. I., G. R. Abecasis, L. R. Cardon, D. B. Goldstein, J. Little, J. P. A. Ioannidis, and J. N. Hirschhorn. 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics* 9:356-369.
- McClellan, A. J., Y. Xia, A. M. Deutschbauer, R. W. Davis, M. Gerstein, and J. Frydman. 2007. Diverse cellular functions of the Hsp90 molecular chaperone uncovered using systems approaches. *Cell* 131:121-135.
- McClintock, B. 1950. The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. U. S. A.* 36:344-355.
- McDonald, J. H. and M. Kreitman. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351:652-654.
- McDonald, M., Y.-Y. Hsieh, Y.-H. Yu, S.-L. Chang, and J.-Y. Leu. 2012. The Evolution of Low Mutation Rates in Experimental Mutator Populations of *Saccharomyces cerevisiae*. *Current biology : CB* 22:1235-1240.
- McDonald, M. J. 2019. Microbial Experimental Evolution - a proving ground for evolutionary theory and a tool for discovery. *EMBO Rep* 20:e46992.
- McKenzie, S. L., S. Henikoff, and M. Meselson. 1975. Localization of RNA from heat-induced polysomes at puff sites in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U. S. A.* 72:1117-1121.
- McLaughlin, S. H., F. Sobott, Z. P. Yao, W. Zhang, P. R. Nielsen, J. G. Grossmann, E. D. Laue, C. V. Robinson, and S. E. Jackson. 2006. The co-chaperone p23 arrests the Hsp90 ATPase cycle to trap client proteins. *J. Mol. Biol.* 356:746-758.
- Melamed, D., D. L. Young, C. E. Gamble, C. R. Miller, and S. Fields. 2013. Deep mutational scanning of an RRM domain of the *Saccharomyces cerevisiae* poly(A)-binding protein. *RNA* 19:1537-1551.
- Mendel, G. J. 1866. "Versuche über Pflanzen-Hybriden" [Experiments Concerning Plant Hybrids]. *Verhandlungen des naturforschenden Vereines in Brünn* [Proceedings of the Natural History Society of Brünn] IV.
- Meselson, M. and F. W. Stahl. 1958. The Replication of DNA in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 44:671-682.
- Meyer, P., C. Prodromou, B. Hu, C. Vaughan, S. M. Roe, B. Panaretou, P. W. Piper, and L. H. Pearl. 2003. Structural and functional analysis of the middle segment of hsp90: implications for ATP hydrolysis and client protein and cochaperone interactions. *Mol. Cell* 11:647-658.
- Meyer, P., C. Prodromou, C. Liao, B. Hu, S. Mark Roe, C. K. Vaughan, I. Vlastic, B. Panaretou, P. W. Piper, and L. H. Pearl. 2004. Structural basis for recruitment of the ATPase activator Aha1 to the Hsp90 chaperone machinery. *The EMBO journal* 23:511-519.
- Millson, S. H., A. W. Truman, V. King, C. Prodromou, L. H. Pearl, and P. W. Piper. 2005. A two-hybrid screen of the yeast proteome for Hsp90 interactors uncovers a novel Hsp90 chaperone

- requirement in the activity of a stress-activated mitogen-activated protein kinase, Slt2p (Mpk1p). *Eukaryotic cell* 4:849-860.
- Mishra, P., J. M. Flynn, T. N. Starr, and D. N. A. Bolon. 2016. Systematic Mutant Analyses Elucidate General and Client-Specific Aspects of Hsp90 Function. *Cell Rep.* 15:588-598.
- Mukai, T. 1964. The Genetic Structure of Natural Population of *Drosophila Melanogaster*. Spontaneous Mutation Rate of Polygenes Controlling Viability. *Genetics* 50:1-19.
- Muller, H. J. 1927. Artificial Transmutation of the Gene *Science* 66:84-87.
- Mumby, P. J. and R. van Woesik. 2014. Consequences of ecological, evolutionary and biogeochemical uncertainty for coral reef responses to climatic stress. *Curr. Biol.* 24:R413-423.
- Mustonen, V. and M. Lässig. 2009. From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation. *Trends Genet.* 25:111-119.
- Mustonen, V. L., M. 2009. From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation. *Trends Genet.* 25:111-119.
- Nardai, G., B. Sass, J. Eber, G. Orosz, and P. Csermely. 2000. Reactive cysteines of the 90-kDa heat shock protein, Hsp90. *Arch. Biochem. Biophys.* 384:59-67.
- Nathan, D. F. and S. Lindquist. 1995. Mutational analysis of Hsp90 function: interactions with a steroid receptor and a protein kinase. *Mol. Cell. Biol.* 15:3917-3925.
- Nathan, D. F., M. H. Vos, and S. Lindquist. 1997. In vivo functions of the *Saccharomyces cerevisiae* Hsp90 chaperone. *Proc. Natl. Acad. Sci. U. S. A.* 94:12949-12956.
- Ng, S. B., K. J. Buckingham, C. Lee, A. W. Bigham, H. K. Tabor, K. M. Dent, C. D. Huff, P. T. Shannon, E. W. Jabs, D. A. Nickerson, J. Shendure, and M. J. Bamshad. 2010. Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.* 42:30-35.
- Nielsen, R. and Z. Yang. 2003. Estimating the distribution of selection coefficients from phylogenetic data with applications to mitochondrial and viral DNA. *Mol. Biol. Evol.* 20:1231-1239.
- Nirenberg, M., P. Leder, M. Bernfield, R. Brimacombe, J. Trupin, F. Rottman, and C. O'Neal. 1965. RNA codewords and protein synthesis, VII. On the general nature of the RNA code. *Proc. Natl. Acad. Sci. U. S. A.* 53:1161-1168.
- Nirenberg, M. W. and J. H. Matthaei. 1961. The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* 47:1588-1602.
- Nollen, E. A. A. and R. I. Morimoto. 2002. Chaperoning signaling pathways: molecular chaperones as stress-sensing 'heat shock' proteins. *J. Cell Sci.* 115:2809-2816.
- Novick, A. and L. Szilard. 1950. Experiments with the Chemostat on spontaneous mutations of bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 36:708-719.
- Ohta, T. 1973. Slightly deleterious mutant substitutions in evolution. *Nature* 246:96-98.
- Ohta, T. 1992. The Nearly Neutral Theory of Molecular Evolution. *Annu. Rev. Ecol. Syst.* 23:263-286.
- Onuoha, S. C., E. T. Coulstock, J. G. Grossmann, and S. E. Jackson. 2008. Structural studies on the co-chaperone Hop and its complexes with Hsp90. *J. Mol. Biol.* 379:732-744.
- Ostrowski, E. A., D. E. Rozen, and R. E. Lenski. 2005. Pleiotropic effects of beneficial mutations in *Escherichia coli*. *Evolution* 59:2343-2352.
- Panaretou, B., C. Prodromou, S. M. Roe, R. O'Brien, J. E. Ladbury, P. W. Piper, and L. H. Pearl. 1998. ATP binding and hydrolysis are essential to the function of the Hsp90 molecular chaperone in vivo. *The EMBO journal* 17:4829-4836.
- Park, S. J., B. N. Borin, M. A. Martinez-Yamout, and H. J. Dyson. 2011. The client protein p53 adopts a molten globule-like state in the presence of Hsp90. *Nat. Struct. Mol. Biol.* 18:537-541.
- Peck, J. R., G. Barreau, and S. C. Heath. 1997. Imperfect genes, Fisherian mutation and the evolution of sex. *Genetics* 145:1171-1199.
- Peisajovich, S. G. and D. S. Tawfik. 2007. Protein engineers turned evolutionists. *Nature methods* 4:991-994.



- Picard, D., B. Khursheed, M. J. Garabedian, M. G. Fortin, S. Lindquist, and K. R. Yamamoto. 1990. Reduced levels of hsp90 compromise steroid receptor action in vivo. *Nature* 348:166-168.
- Piganeau, G. and A. Eyre-Walker. 2003. Estimating the distribution of fitness effects from DNA sequence data: implications for the molecular clock. *Proc. Natl. Acad. Sci. U. S. A.* 100:10335-10340.
- Piper, P. W. 1995. The heat shock and ethanol stress responses of yeast exhibit extensive similarity and functional overlap. *FEMS Microbiol. Lett.* 134:121-127.
- Plotkin, J. B. and G. Kudla. 2011. Synonymous but not the same: the causes and consequences of codon bias. *Nature reviews. Genetics* 12:32-42.
- Poelwijk, F. J., D. J. Kiviet, D. M. Weinreich, and S. J. Tans. 2007. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* 445:383-386.
- Posas, F., J. R. Chambers, J. A. Heyman, J. P. Hoeffler, E. de Nadal, and J. Arino. 2000. The transcriptional response of yeast to saline stress. *J. Biol. Chem.* 275:17249-17255.
- Prodromou, C. 2016. Mechanisms of Hsp90 regulation. *The Biochemical journal* 473:2439-2452.
- Quail, M. A., M. Smith, P. Coupland, T. D. Otto, S. R. Harris, T. R. Connor, A. Bertoni, H. P. Swerdlow, and Y. Gu. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 13:341.
- Queitsch, C., T. A. Sangster, and S. Lindquist. 2002. Hsp90 as a capacitor of phenotypic variation. *Nature* 417:618-624.
- Ratcliff, W. C., R. F. Denison, M. Borrello, and M. Travisano. 2012. Experimental evolution of multicellularity. *Proceedings of the National Academy of Sciences* 109:1595-1600.
- Retzlaff, M., M. Stahl, H. C. Eberl, S. Lagleder, J. Beck, H. Kessler, and J. Buchner. 2009. Hsp90 is regulated by a switch point in the C-terminal domain. *EMBO reports* 10:1147-1153.
- Revie, D., D. W. Smith, and T. W. Yee. 1988. Kinetic analysis for optimization of DNA ligation reactions. *Nucleic Acids Res.* 16:10301-10321.
- Reznick, D. N. and C. K. Ghalambor. 2005. Selection in Nature: Experimental Manipulations of Natural Populations1. *Integr. Comp. Biol.* 45:456-462.
- Richter, K., M. Haslbeck, and J. Buchner. 2010. The heat shock response: life on the verge of death. *Mol. Cell* 40:253-266.
- Roe, S. M., M. M. Ali, P. Meyer, C. K. Vaughan, B. Panaretou, P. W. Piper, C. Prodromou, and L. H. Pearl. 2004. The Mechanism of Hsp90 regulation by the protein kinase-specific cochaperone p50(cdc37). *Cell* 116:87-98.
- Rohl, A., J. Rohrberg, and J. Buchner. 2013. The chaperone Hsp90: changing partners for demanding clients. *Trends Biochem. Sci.* 38:253-262.
- Rohner, N., D. F. Jarosz, J. E. Kowalko, M. Yoshizawa, W. R. Jeffery, R. L. Borowsky, S. Lindquist, and C. J. Tabin. 2013. Cryptic Variation in Morphological Evolution: HSP90 as a Capacitor for Loss of Eyes in Cavefish. *Science* 342:1372-1375.
- Roscoe, B. P. and D. N. A. Bolon. 2014. Systematic exploration of ubiquitin sequence, E1 activation efficiency, and experimental fitness in yeast. *J. Mol. Biol.* 426:2854-2870.
- Roscoe, B. P., K. M. Thayer, K. B. Zeldovich, D. Fushman, and D. N. A. Bolon. 2013. Analyses of the effects of all ubiquitin point mutants on yeast growth rate. *J. Mol. Biol.* 425:1363-1377.
- Rozen, D. E., L. McGee, B. R. Levin, and K. P. Klugman. 2007. Fitness costs of fluoroquinolone resistance in *Streptococcus pneumoniae*. *Antimicrob. Agents Chemother.* 51:412-416.
- Rutherford, S., Y. Hirate, and B. J. Swalla. 2007. The Hsp90 capacitor, developmental remodeling, and evolution: the robustness of gene networks and the curious evolvability of metamorphosis. *Crit. Rev. Biochem. Mol. Biol.* 42:355-372.
- Rutherford, S. L. 2003. Between genotype and phenotype: protein chaperones and evolvability. *Nature reviews. Genetics* 4:263-274.

- Rutherford, S. L. and S. Lindquist. 1998. Hsp90 as a capacitor for morphological evolution. *Nature* 396:336-342.
- Ruvkun, G. B. and F. M. Ausubel. 1981. A general method for site-directed mutagenesis in prokaryotes. *Nature* 289:85-88.
- Saito, H. and F. Posas. 2012. Response to hyperosmotic stress. *Genetics* 192:289-318.
- Sane, M., J. J. Miranda, and D. Agashe. 2018. Antagonistic pleiotropy for carbon use is rare in new mutations. *Evolution* 72:2202-2213.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* 74:5463-5467.
- Sangster, T. A., S. Lindquist, and C. Queitsch. 2004. Under cover: causes, effects and implications of Hsp90-mediated genetic capacitance. *Bioessays* 26:348-362.
- Sanjuán, R., A. Moya, and S. F. Elena. 2004. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc. Natl. Acad. Sci. U. S. A.* 101:8396-8401.
- Sarkisyan, K. S., D. A. Bolotin, M. V. Meer, D. R. Usmanova, A. S. Mishin, G. V. Sharonov, D. N. Ivankov, N. G. Bozhanova, M. S. Baranov, O. Soylemez, N. S. Bogatyreva, P. K. Vlasov, E. S. Egorov, M. D. Logacheva, A. S. Kondrashov, D. M. Chudakov, E. V. Putintseva, I. Z. Mamedov, D. S. Tawfik, K. A. Lukyanov, and F. A. Kondrashov. 2016. Local fitness landscape of the green fluorescent protein. *Nature* 533:397-401.
- Sato, S., N. Fujita, and T. Tsuruo. 2000. Modulation of Akt kinase activity by binding to Hsp90. *Proc. Natl. Acad. Sci. U. S. A.* 97:10832-10837.
- Schneider, A., B. Charlesworth, A. Eyre-Walker, and P. D. Keightley. 2011. A method for inferring the rate of occurrence and fitness effects of advantageous mutations. *Genetics* 189:1427-1437.
- Schochetman, G., C. Y. Ou, and W. K. Jones. 1988. Polymerase chain reaction. *J. Infect. Dis.* 158:1154-1157.
- Schopf, F. H., M. M. Biebl, and J. Buchner. 2017. The HSP90 chaperone machinery. *Nature Reviews Molecular Cell Biology* 18:345-360.
- Silander, O. K., O. Tenaillon, and L. Chao. 2007. Understanding the evolutionary fate of finite populations: the dynamics of mutational effects. *PLoS Biol.* 5:e94.
- Smith, J. M. and J. Haigh. 1974. The hitch-hiking effect of a favourable gene. *Genet. Res.* 23:23-35.
- Smith, M. M., L. Shi, and M. Navre. 1995. Rapid Identification of Highly Active and Selective Substrates for Stromelysin and Matrilysin Using Bacteriophage Peptide Display Libraries. *J. Biol. Chem.* 270:6440-6449.
- Sniegowski, P. D. and P. J. Gerrish. 2010. Beneficial mutations and the dynamics of adaptation in asexual populations. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 365:1255-1263.
- Soskine, M. and D. S. Tawfik. 2010. Mutational effects and the evolution of new protein functions. *Nature Reviews Genetics* 11:572-582.
- Sousa, A., S. Magalhaes, and I. Gordo. 2012. Cost of antibiotic resistance and the geometry of adaptation. *Mol. Biol. Evol.* 29:1417-1428.
- Southworth, D. R. and D. A. Agard. 2011. Client-loading conformation of the Hsp90 molecular chaperone revealed in the cryo-EM structure of the human Hsp90:Hop complex. *Mol. Cell* 42:771-781.
- Stankiewicz, M. and M. P. Mayer. 2012. The universe of Hsp90. *Biomol. Concepts* 3:79-97.
- Starita, L. M., J. N. Pruneda, R. S. Lo, D. M. Fowler, H. J. Kim, J. B. Hiatt, J. Shendure, P. S. Brzovic, S. Fields, and R. E. Klevit. 2013. Activity-enhancing mutations in an E3 ubiquitin ligase identified by high-throughput mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.* 110:E1263-1272.
- Starr, T. N., J. M. Flynn, P. Mishra, D. N. A. Bolon, and J. W. Thornton. 2018. Pervasive contingency and entrenchment in a billion years of Hsp90 evolution. *Proceedings of the National Academy of Sciences* 115:4453-4458.

- Steinberg, B. and M. Ostermeier. 2016. Environmental changes bridge evolutionary valleys. *Sci Adv* 2:e1500921-e1500921.
- Stemmer, W. P. 1994. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* 370:389-391.
- Stiffler, M., D. Hekstra, and R. Ranganathan. 2015. Evolvability as a Function of Purifying Selection in TEM-1  $\beta$ -Lactamase. *Cell* 160:882-892.
- Taipale, M., D. F. Jarosz, and S. Lindquist. 2010. HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nat. Rev. Mol. Cell Biol.* 11:515-528.
- Taipale, M., I. Krykbaeva, M. Koeva, C. Kayatekin, K. D. Westover, G. I. Karras, and S. Lindquist. 2012. Quantitative analysis of HSP90-client interactions reveals principles of substrate recognition. *Cell* 150:987-1001.
- Team., R. C. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Tenaillon, O. 2014. The Utility of Fisher's Geometric Model in Evolutionary Genetics. *Annu. Rev. Ecol. Evol. Syst.* 45:179-201.
- Thatcher, J. W., J. M. Shaw, and W. J. Dickinson. 1998. Marginal fitness contributions of nonessential genes in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 95:253-257.
- Tschermak, E. 1950. Concerning Artificial Crossing in *Pisum Sativum*. *Genetics* 35.
- Tuller, T., A. Carmi, K. Vestsigian, S. Navon, Y. Dorfan, J. Zaborske, T. Pan, O. Dahan, I. Furman, and Y. Pilpel. 2010. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 141:344-354.
- Tutt, J. 1896. *British Moths*. London:Routledge.
- Venkataram, S., B. Dunn, Y. Li, A. Agarwala, J. Chang, E. R. Ebel, K. Geiler-Samerotte, L. Hérissant, J. R. Blundell, S. F. Levy, D. S. Fisher, G. Sherlock, and D. A. Petrov. 2016. Development of a Comprehensive Genotype-to-Fitness Map of Adaptation-Driving Mutations in Yeast. *Cell* 166:1585-1596.e1522.
- Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferreira, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y. H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigó, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H.

- Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. H. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh and X. Zhu. 2001. The sequence of the human genome. *Science* 291:1304-1351.
- Verba, K. A., R. Y. Wang, A. Arakawa, Y. Liu, M. Shirouzu, S. Yokoyama, and D. A. Agard. 2016. Atomic structure of Hsp90-Cdc37-Cdk4 reveals that Hsp90 traps and stabilizes an unfolded kinase. *Science* 352:1542-1547.
- Wallace, A. R. 2013. Natural History Museum: Wallace Letters Online.
- Watson, J. D. and F. H. C. Crick. 1953. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* 171:737-738.
- Waxman, D. 2006. Fisher's geometrical model of evolutionary adaptation—Beyond spherical geometry. *J. Theor. Biol.* 241:887-895.
- Wayne, N. and D. N. Bolon. 2007. Dimerization of Hsp90 is required for in vivo function. Design and analysis of monomers and dimers. *J. Biol. Chem.* 282:35386-35395.
- Weinreich, D. M., N. F. Delaney, M. A. DePristo, and D. L. Hartl. 2006. Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. *Science* 312:111-114.
- Weiss, G. A., C. K. Watanabe, A. Zhong, A. Goddard, and S. S. Sidhu. 2000. Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc. Natl. Acad. Sci. U. S. A.* 97:8950-8954.
- Wells, J. A., M. Vasser, and D. B. Powers. 1985. Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites. *Gene* 34:315-323.
- Whitehead, T. A., A. Chevalier, Y. Song, C. Dreyfus, S. J. Fleishman, C. De Mattos, C. A. Myers, H. Kamisetty, P. Blair, I. A. Wilson, and D. Baker. 2012. Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat. Biotechnol.* 30:543-548.
- Whitesell, L., E. G. Mimnaugh, B. De Costa, C. E. Myers, and L. M. Neckers. 1994. Inhibition of heat shock protein HSP90-pp60v-src heteroprotein complex formation by benzoquinone ansamycins: essential role for stress proteins in oncogenic transformation. *Proc. Natl. Acad. Sci. U. S. A.* 91:8324-8328.
- Wielgoss, S., J. E. Barrick, O. Tenaillon, M. J. Wisner, W. J. Dittmar, S. Cruveiller, B. Chane-Woon-Ming, C. Médigue, R. E. Lenski, and D. Schneider. 2013. Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proceedings of the National Academy of Sciences* 110:222-227.
- Wisner, M. J., N. Ribbeck, and R. E. Lenski. 2013. Long-Term Dynamics of Adaptation in Asexual Populations. *Science* 342:1364-1367.
- Wloch, D. M., K. Szafraniec, R. H. Borts, and R. Korona. 2001. Direct estimate of the mutation rate and the distribution of fitness effects in the yeast *Saccharomyces cerevisiae*. *Genetics* 159:441-452.
- Wolmarans, A., B. Lee, L. Spyropoulos, and P. LaPointe. 2016. The Mechanism of Hsp90 ATPase Stimulation by Aha1. *Sci. Rep.* 6:33179.
- Wright, S. 1932. The Roles of Mutation, Inbreeding, Crossbreeding, and Selection in Evolution. *Proceedings of the Sixth International Congress of Genetics*:356-366.

- Wu, N. C., C. A. Olson, Y. Du, S. Le, K. Tran, R. Remenyi, D. Gong, L. Q. Al-Mawsawi, H. Qi, T.-T. Wu, and R. Sun. 2015. Functional Constraint Profiling of a Viral Protein Reveals Discordance of Evolutionary Conservation and Functionality. *PLoS Genet.* 11:e1005310-e1005310.
- Wu, N. C., A. P. Young, S. Dandekar, H. Wijersuriya, L. Q. Al-Mawsawi, T.-T. Wu, and R. Sun. 2013. Systematic identification of H274Y compensatory mutations in influenza A virus neuraminidase by high-throughput screening. *J. Virol.* 87:1193-1199.
- Wu, R. 1970. Nucleotide sequence analysis of DNA. I. Partial sequence of the cohesive ends of bacteriophage lambda and 186 DNA. *J. Mol. Biol.* 51:501-521.
- Wu, R. and A. D. Kaiser. 1968. Structure and base sequence in the cohesive ends of bacteriophage lambda DNA. *J. Mol. Biol.* 35:523-537.
- Xu, Y. and S. Lindquist. 1993. Heat-shock protein hsp90 governs the activity of pp60v-src kinase. *Proc. Natl. Acad. Sci. U. S. A.* 90:7074-7078.
- Yang, X. X., P. Hawle, J. P. Bebelman, A. Meenhuis, M. Siderius, and S. M. van der Vies. 2007. Cdc37p is involved in osmoadaptation and controls high osmolarity-induced cross-talk via the MAP kinase Kss1p. *FEMS yeast research* 7:796-807.
- Yang, X. X., K. C. Maurer, M. Molanus, W. H. Mager, M. Siderius, and S. M. van der Vies. 2006. The molecular chaperone Hsp90 is required for high osmotic stress response in *Saccharomyces cerevisiae*. *FEMS yeast research* 6:195-204.
- Zhang, M., Y. Kadota, C. Prodromou, K. Shirasu, and L. H. Pearl. 2010. Structural basis for assembly of Hsp90-Sgt1-CHORD protein complexes: implications for chaperoning of NLR innate immunity receptors. *Mol. Cell* 39:269-281.
- Zhang, R., D. Luo, R. Miao, L. Bai, Q. Ge, W. C. Sessa, and W. Min. 2005. Hsp90–Akt phosphorylates ASK1 and inhibits ASK1-mediated apoptosis. *Oncogene* 24:3954-3963.
- Zhao, H. and F. H. Arnold. 1999. Directed evolution converts subtilisin E into a functional equivalent of thermitase. *Protein Eng.* 12:47-53.
- Zirkle, C. 1941. Natural Selection before the "Origin of Species". *Proceedings of the American Philosophical Society* 84:71-123.
- Zuckerlandl, E. and L. Pauling. 1965. Molecules as documents of evolutionary history. *J. Theor. Biol.* 8:357-366.