安徽农业大学学报,2005,32(3):309~315 Joumal of Anhui Agricultural University

The Phylogeny Tree Reconstruction Based on the Usage Frequency of Codons and Corresponding Complementary Codons

HUANG Hui-fang¹, ZHUANG Yong-long², CHEN Liming¹, LN Luping¹, MA Fei^{3*}, XU Xiao-feng⁴ (1. School of Life Science, Xianen University, Xianen 361005; 2. Institute of Bioinformatics, Tsinghua University, Beijing 100084; 3. Laboratory of Comparative Genome and Bioinformatics, College of Life Science, Liaoning Normal University, Dalian 116029; 4. Life Science College, Nanjing Normal University, Nanjing 210097)

Abstract: The hypothesis that a codon bias usage is identical to its complementary codon usage preference has been investigated by using the relationship analysis of codons vs their complementary ones among 70 organisms. Significantly positive usage correlations between codons and their complementary ones were found and its implication in biology was also analyzed. The codon-complementary codon tree was further built, which fairly exhibited the evolutionary relationship of these organisms. The results not only demonstrated the validity of our hypothesis, but also manifested the usefulness of correlation analysis in studying on codon usage pattern and molecular evolutionary mechanisms of organism.

Key words: codon; complementary codon; usage frequency; usage bias; correlation analysis; codon-complementary codon tree

CLC num ber:Q987 Docum ent code:A Article D:1672-352X (2005) 03-0309-07

Genetic code is one of the most important biological languages in communications between DNA and protein. Revealing the origin and evolution mechanism of this language will bring an invaluable significance to revalue the mystery of life origin The focus is now on what is the driving force of this selection and how does the selection work As the logic of origination and evolution of genetic code may be staying hidden in this kind of driving force and selection mechanism, many exciting researches in the right area have been made^[1~9]. Grosiean and co-worker^[1,2] put forward a hypothesis namely "the optimum energy of codon-anticodon interaction" for explaining codon usage bias Ikemura^[3~6] found that the usage bias of a codon is usually determined by the abundance of its corresponding tRNAs in some unicellular organisms including E. coli and yeast, known as "Ikemura rule". The hypothesis of translation efficiency was used for explaining codons usage bias As a matter of fact, we found that these two hypotheses, both being based on the interaction of codons with their corresponding anticodons, are essentially the same. If these hypotheses hold true, the optimum energy of codon-anticodon interaction should be the essential prerequisite of translation efficiency of a protein, a positive correlation of codons with their complementary codons in usage should exist This is because that the pair match energy of any codon with its anticodon is equal to that of this codon s complementary codon with its own anticodon Let s take codons 'CCG' and 'CCA' as examples, 'CCG' s anticodon and complementary codon are 'CGG' and 'CGG', the same nucleotide triplet, and 'CCA' s anticodon and complementary codon 'UGG' and 'UGG', also the same kind of nucleotide triplet It is obvious that the complexes of codon-anticodon equal to the complexes of corresponding complementary codon-codon, such as 'CCG-CGG 'vs 'CGG-CCG ' and 'CCA-UGG 'vs 'UGG-CCA ' in their

Biography: * Corresponding author E-mail: mafei01@mail. tsinghua.org. cn

Received date: 2004-02-16

Foundation item: This work was supported by a grant from the National Natural Science Foundation of China (No. 60305001) and a grant from the Science and Technology Ministry of China (Preliminary Study on Functional Genome Systematics) (No. 2001CCA01400).

combination energy. Therefore, if the pair-matching energy is the main selection force of a codon s usage, all corrsponding complementary codons should possess the same usage pattern as their corresponding codons. In other words, any complementary codon should have a strictly positive correlation with its codon in their usage frequency. If so, the pair-matching energy of a codon with its anticodon should probably be a selection force for the codons usage, which just as Grosjean and Ikemura ever proposed. Hence, we put forward the hypothesis, namely when a codon is biased in usage, its complementary codon is also preferred in an organism. In order to testify the hypothesis and figure out what kind of usage relationship between codons and their complementary ones from the genomes of 70 organisms including 45 Bacteria, 5 A rchaea, and 20 Eukarya in this paper, and found that there in deed exists a positive usage correlation relationship between codons and their complementary ones

1 Materials and methods

Genomic data of 70 organisms: 5 A rchaea $(1 \sim 5)$, 20 Eukaryota $(6 \sim 25)$ and 45 B acteria $(26 \sim 70)$ in Table 1 were obtained from the Kazusa DNA Research Institute (http://www. kazusa.or. jp/codon/, GenB ank Release 129.0, 15 April 2002). Usage correlation between all codons and their complementary ones of 70 organisms were studied using correlation analysis. The correlation coefficients between 58 codons with their corresponding complementary codons (three terminal codons and their complementary ones were excluded) in usage were calculated by using the usage frequency data of 58 codons as the independent variables and those of their corresponding complementary codons as dependent variables for all 70 organisms

O rganism s	Domains		No. of codons	GC content	GC3 s content	r ₅₈
01. M. jannaschii	A rchaea	Themphile	504 594	31.85	24.73	0. 3254*
02. M. them autotrophicus		Themphile	595 693	50.44	56.54	0.3903**
03. P. horikoshii		Themphile	486 103	42.45	43.37	0.3728 ^{**}
04. T. acidophilum		Themphile	471 498	47.38	55.04	0.2914*
05. T. volcanium		Themphile	454 874	40.99	40.71	0.4027 * *
06. Hom o sapiens	Eukaryota	Vertebrate	19 894 411	52.65	59.31	0. 4940 * *
07. Sus scrofa		Vertebrate	406 321	54.07	64.80	0.4804 * *
08. O. cuniculus		Vertebrate	443 600	54.55	67.27	0.4503 * *
09. M us m usculus		Vertebrate	9 549 215	52.41	59.45	0. 4981 * *
10. Rattus norvegicus		Vertebrate	3 435 705	52.79	61.26	0.4635 ^{**}
11. X. laevis		Vertebrate	973 234	47.35	48.81	0.3145*
12. C. elegans		Pro to stome	9 684 274	42.76	39.89	0.4766 ^{**}
13. S. cerevisiae		Fungi	5 664 727	39.70	37.94	0.4790**
14. K. lactis		Fungi	114 708	39.09	35.88	0.4485 * *
15. C. albicans		Fungi	330 853	36.90	28.96	0.4592**
16. <i>S. pan be</i>		Fungi	2 840 951	39.80	33.12	0.3693**
17. N. crassa		Fungi	939 373	56.21	65.58	0.5062**
18. E. nidulans		Fungi	204 431	53.17	58.86	0.5057**
19. A. niger		Fungi	82 096	56.22	67.84	0.4781 * *
20. A. thaliana		Plant	19 602 801	44.44	42.19	0.3982**
21. O. sativa		Plant	1 701 592	54.60	61.26	0. 4974 * *
22. N. tabacum		Plant	347 019	43.55	39.49	0. 2965 *
23. Zea m ays		Plant	556 901	54.88	63.94	0.4613**
24. H. vulgare subsp. vulgare		Plant	151 278	58.06	71.60	0. 5089 * *
25. L. esculen tum		Plant	391 001	42.62	37.61	0.3344*
26. E. coli K12	Bacteria	subdivision	1 363 716	51.83	55.89	0.4913**

Table 1The usage correlation of codons vs complementary codons of the genomes in 70 organisms

Table 1 continued

27. E. coli (Gam-negative protocolor:in) subdivision 3 662 594 50. 58 53. 6. 0. 4716 ¹¹ 28. S. pylin ariam LT2 subdivision 177 278 53. 6. 59. 30 0. 5465 ¹ 30. K. purcam oriate subdivision 178 416 55. 77 64. 84 0. 6350 ¹¹ 31. Y. enterncolitica subdivision 171 17 48. 97 50. 17 0. 4963 ¹¹ 32. Y. petita subdivision 1399 331 47. 35 47. 69 0. 4346 ¹¹ 33. V. cholerae subdivision 1399 331 47. 35 47. 69 0. 4346 ¹¹ 34. II. influenze Rd subdivision 86 514 42. 11 30. 68 0. 3975 ¹¹ 35. C. humenija subdivision 1342 66. 44 86. 27 0. 6446 ¹¹ 39. P. putida subdivision 1342 66. 44 86. 27 0. 6447 ¹¹ 41. B. pertasi subdivision 1343 65. 24 85. 99 0. 6167 ¹¹ 42. N. ne metingitidis subdivision 303 391 51.12 56. 80 90. 6167 ¹¹ 43. N. ne metingitidis subdivision 303 391							
29. S. typhin urium LT2 subdivision 1 477 278 53. 36 59. 53 0. 5465 ⁺ 30. K. pream mine subdivision 178 416 55. 77 64. 84 0. 6530 ⁺ 31. Y. enteracolitica subdivision 123 436 47. 17 44. 54 0. 0496 ⁺ 32. Y. pestis subdivision 129 331 47. 35 47. 69 0. 4346 ⁺ 33. V. choleme subdivision 523 322 38. 76 29.08 0. 2800 ⁺ 35. C. burnetii subdivision 98 514 42. 51 39.68 0. 5578 ⁺ 36. B. aphidicoln subdivision 90 279 26. 99 13. 66 0. 5578 ⁺ 37. Buchnem up, APS subdivision 128 858 27. 43 14. 29 0. 5831 ⁺ 38. P. aerginson subdivision 78 438 65.24 88. 93 0. 6167 ⁺ 41. R. pernatia subdivision 78 438 65.24 89.30 0. 6675 ⁺ 42. N. ameringtidis subdivision 78 438 0.65.24 89.04 0.5785 ⁺ 43. N. meringtidis subdivision 78 438 0.65.24 89.04 0.5	27. E. coli	(Gram-negative	subdivision	3 662 594	50.58	53.36	0.4716 ^{**}
30. K. meam.oniae subdivision 178 416 55. 77 64. 84 0.6550' 31. Y. enterncolitica subdivision 123 436 47. 17 44. 50 0.4056' 32. Y. choeme subdivision 1471 174 48. 97 50. 17 0.4366' 33. V. choeme subdivision 123 436 47. 17 44. 87 0.4346' 34. H. influenzae Rd subdivision 923 322 38. 76 0.2306 0.2300' 56. E. aphiticola subdivision 90 279 26. 99 13. 66 0.578'' 37. Buchnera sp. APS subdivision 188 858 27. 43 14. 29 0.5831'' 38. P. aeraginosa subdivision 78 438 65.24 85.93 0.6167'' 40. A. vinetandii subdivision 78 438 65.24 85.93 0.6167'' 41. B. pertosis subdivision 78 438 65.24 85.93 0.6167'' 42. N. meningitidis subdivision 78 438 65.24 85.93 0.6167'' 43. N. meningitidis subdivision 859 048 53.06 61.39 0.7757''	28. S. typhim u rium	proteobacteria)	subdivision	360 681	52.82	58.70	0.5345**
31. Y. entencollica subdivision 12.3 4.36 47.17 46.45 0.4056 ⁺ 32. Y. positi subdivision 1.471 174 48.97 50.17 0.4963 ⁺ 33. V. choleme subdivision 1.399 331 47.35 47.69 0.4346 ⁺ 34. H. influenze Rd subdivision 8.25 22.2 2.37.6 2.080 0.2800 ⁺ 35. C. bu metii subdivision 90 279 26.99 13.66 0.5578 ⁺ 36. B. aphidicola subdivision 90 279 26.99 13.66 0.5578 ⁺ 37. Ruchnerus p. APS subdivision 12.3142 66.44 86.27 0.6446 ⁺ 39. P. patida subdivision 91.147 67.87 85.13 0.1677 ⁺ 41. B. pertussit subdivision 91.147 67.87 85.13 0.1677 ⁺ 42. N. genorthocae subdivision 140 675 52.56 69.29 0.5768 ⁺ 43. N. mening itidis subdivision 168 270 59.71 71.48 0.5768 ⁺ 44. N. m. ening itidis subdivision 168 270 59.77 71.48 0.7561 ⁺⁺ <td>29. S. typh in u rium LT2</td> <td></td> <td>subdivision</td> <td>1 477 278</td> <td>53.36</td> <td>59.53</td> <td>0. 5465 * *</td>	29. S. typh in u rium LT2		subdivision	1 477 278	53.36	59.53	0. 5465 * *
32. Y. pestis subdivision 1 471 174 48.97 50. 17 0.4963 ⁻¹ 33. V. choleme subdivision 1 39 331 47.35 47.69 0.4346 ⁻¹ 34. H. influenzae Rd subdivision 523 322 38.76 29.08 0.2800 ⁻¹ 35. C. humetii subdivision 86 514 42.51 39.68 0.2893 ⁻¹ 36. B. aphidicola subdivision 188 858 27.43 14.29 0.5831 ⁻¹ 38. P. aengi pinoa subdivision 213 342 66.24 86.27 0.6446 ⁻¹ 39. P. putida subdivision 78 438 65.24 85.93 0.1617 ⁻¹ 41. B. pertussis subdivision 19147 67.87 85.13 0.7175 ⁻¹ 42. N. gonorrhoeze subdivision 20147 56.26 59.92 0.5960 ⁻¹ 43. N. menigitidit X240 subdivision 530 391 51.12 56.80 0.5768 ⁻¹ 45. N. me nigitidit X240 subdivision 583 891 53.32 62.08 0.5660 ⁻¹ 45. N. me nigitidit X2401 subdivision 590 48 53.64 0.7971 ⁻¹ 71.48 </td <td>30. K. pneum on iae</td> <td></td> <td>subdivision</td> <td>178 416</td> <td>55.77</td> <td>64.84</td> <td>0.6530**</td>	30. K. pneum on iae		subdivision	178 416	55.77	64.84	0.6530**
33. V. choleme subdivision 1 399 331 47. 35 47. 69 0. 4346 ⁺ . 34. H. influenzae Rd subdivision 523 322 38. 76 29. 08 0. 2800 ⁺ 35. C. humetti subdivision 86 514 42. 51 39. 68 0.3975 ⁺ 36. B. aphilicola subdivision 90 79 26. 99 13. 66 0.5578 ⁺ 37. Bucknern sp. APS subdivision 2313 442 66. 44 86. 27 0.6446 ⁺ 39. P. putida subdivision 3145 718 60. 25 73. 64 0.6373 ⁺ 41. B. pertassis subdivision 91 147 67.87 85.13 0.7175 ⁺ 42. N. gonorthocae subdivision 303 391 51.12 56.80 0.5768 ⁺ 43. N. me migritidis subdivision 589 048 53.06 61.39 0.5735 ⁺ 44. N. me migritidis subdivision 168 270 59.77 71.48 0.7568 ⁺ 47. A. tm officiens str. C58 subdivision 168 37 60.13 72.71 0.7551 ⁺ 48. A. tam officiens str. C58 subdivision 169312 59.77 71.48	31. Y. enterocolitica		subdivision	123 436	47.17	46.45	0.4056**
34. H. influenzae Rd subdivision 523 322 38. 76 29. 08 0. 2800' 35. C. humetti subdivision 86 514 42. 51 39. 68 0. 3575' 36. B. aphtilicola subdivision 90 279 26. 99 13. 66 0. 5578'' 37. Buchnear ap. APS subdivision 188 858 27. 43 14. 20 0.5331'' 38. P. aeruginosa subdivision 2313 442 66. 44 86. 27 0. 6446'' 39. P. patida subdivision 78. 438 65. 24 85. 93 0. 6167'' 41. B. pertassis subdivision 78. 438 65. 24 85. 93 0. 6167'' 42. N. genornhocae subdivision 580 448 53. 06 61. 39 0.5785'' 44. N. me mignitidi XC28 subdivision 580 448 53. 06 61. 39 0.5785'' 45. N. me faciens str. C58 subdivision 166 770 57. 77 71. 48 0.7656'' 46. A. num efaciens str. C58 subdivision 166 771 57. 60 66. 18 0.8180'' 50. R. hicopinen subdivision 168 771 57. 60 66. 18	32. Y. pestis		subdivision	1 471 174	48.97	50.17	0.4963**
35. C. burnettii subdivision 86 514 42. 51 39. 68 0. 3975 ⁺ 36. B. aphidicola subdivision 90 279 26. 99 13. 66 0. 5578 ⁺ 37. Buchnera sp. APS subdivision 188 858 27. 43 14. 29 0. 5578 ⁺ 38. P. aenginosa subdivision 343 42 66. 44 86. 27 0. 6446 ⁺ 39. P. putida subdivision 345 718 60. 25 73. 64 0. 6373 ⁺ 40. A. vinelandii subdivision 78. 438 65. 24 85. 93 0. 6167 ⁺ 41. B. perussis subdivision 78. 438 65. 24 85. 93 0. 6167 ⁺ 42. N. gonorhoce subdivision 303 391 51. 12 56. 80 0. 5785 ⁺ 43. N. meningitidis subdivision 589 048 53. 06 61. 39 0. 5735 ⁺ 44. N. meningitidis subdivision 168 270 59. 77 71. 48 0. 6768 ⁺ (U.W ashingpin) subdivision 1697 312 59. 77 71. 48 0. 7581 ⁺ 98. I. Regun inscanne subdivision 168 8270 58. 26 68. 24 <	33. V. cholerae		subdivision	1 399 331	47.35	47.69	0.4346**
36. B. aphidicola subdivision 90 279 26.99 13.66 0.5578 ⁺ 37. Buchnern sp. APS subdivision 188 858 27.43 14.29 0.5831 ⁺ 38. P. aernginosa subdivision 213 442 66.44 86.27 0.6446 ⁺ 39. P. putida subdivision 78 438 65.24 85.93 0.6167 ⁺ 40. A. vinelandii subdivision 78 438 65.24 85.13 0.7175 ⁺ 41. B. perussis subdivision 303 391 51.12 56.80 0.5768 ⁺ 42. N. genorthocae subdivision 303 391 51.12 56.80 0.5768 ⁺ 43. N. meningitidis subdivision 583 889 53.32 62.08 0.5660 ⁺ 44. N. ne mingitidis subdivision 1668 270 59.77 71.48 0.7585 ⁺ 45. A. tun efaciens str. C58 subdivision 166371 59.74 71.45 0.7581 ⁺ 49. R. legum inosanen subdivision 166771 59.77 71.48 0.7585 ⁺ 51. R. hizogines subdivision 163774 66.38 83.81 0.581 ⁺ <	34. H. influenzae Rd		subdivision	523 322	38.76	29.08	0.2800^{*}
37. Buchnern sp. APS subdivision 188 858 27.43 14.29 0.5831 ⁺ 38. P. aernginosa subdivision 2 313 442 66.44 86.27 0.6446 ⁺ 39. P. patida subdivision 2 345 718 60.22 73.64 0.6373 ⁺ 41. B. permasis subdivision 78.483 65.24 85.93 0.6167 ⁺ 42. N. gonorrhoeae subdivision 140 675 52.56 59.92 0.5963 ⁺ 43. N. meningitidis subdivision 583 889 53.32 62.08 0.5660 ⁺ 44. N. meningitidis subdivision 294 124 56.72 64.20 0.7771 ⁺ 47. A. tam efaciens stt: C58 subdivision 1 668 270 59.77 71.48 0.7658 ⁺ (U.W ashington) subdivision 1 673 112 59.74 71.45 0.7581 ⁺ 48. A. tam efaciens stt: C58 subdivision 1 668 270 59.77 71.48 0.7659 ⁺ 51. Rhizobium sp. NGR234 subdivision 1 673 112 59.74 71.45 0.7581 ⁺ 52. R. caputatas subdivision 1 677 12 59.77 71.48	35. C. bumetii		subdivision	86 514	42.51	39.68	0.3975**
38. P. aeruginosa subdivision 2 313 442 66.44 86.27 0.6446 ⁺⁺ 39. P. putida subdivision 345 718 60.25 73.64 0.6373 ⁺⁺ 40. A. vinelandii subdivision 78.438 65.24 85.93 0.6167 ⁺⁺ 41. B. pertussis subdivision 91.147 67.87 85.13 0.7175 ⁺⁺ 42. N. genorrheeae subdivision 90.675 52.56 59.92 0.5663 ⁺⁺ 43. N. meningitidis subdivision 303 391 51.12 56.80 0.5768 ⁺⁺ 44. N. meningitidis subdivision 16075 52.56 69.92 0.5763 ⁺⁺ 45. N. meningitidis subdivision 583 889 53.06 61.39 0.5763 ⁺⁺ 46. A. ame faciens subdivision 168270 59.77 71.48 0.7658 ⁺⁺ (U.W ashington) subdivision 166351 57.60 66.18 0.8180 ⁺⁺ 50. R. hicogenes subdivision 163774 66.38 83.81 0.581 ⁺⁺ 51. Rhicohiam sp. NGR234 subdivision 129 817 58.92 68.24 0.7993 ⁺⁺	36. B. aphidicola		subdivision	90 279	26.99	13.66	0.5578^{**}
39. P. putida subdivision 345 718 60.25 73.64 0.6373 ⁺ 40. A. vinclandii subdivision 78 438 65.24 85.93 0.6167 ⁺ 41. B. pertussis subdivision 91 147 67.87 88.13 0.7175 ⁺ 42. N. gonorrhoeae subdivision 91 147 67.87 88.13 0.7175 ⁺ 42. N. gonorrhoeae subdivision 140 675 52.66 59.92 0.5968 ⁺ 43. N. meningitidis subdivision 589 048 53.06 61.39 0.5735 ⁺ 45. N. meningitidis subdivision 294 124 56.72 64.20 0.7771 ⁺ 47. A. tum efaciens subdivision 168 270 59.77 71.48 0.7688 ⁺ (U.W ashington) subdivision 1697 312 59.74 71.45 0.7581 ⁺ 48. A. tum efaciens str. C58 subdivision 163 377 66.18 0.8180 ⁺ 51. R. hizobian sp. NGR234 subdivision 163 774 66.38 83.81 0.5871 ⁺ 52. R. conorri subdivision 312 081 30.63 18.43 0.522 ⁺	37. Buchnera sp. APS		subdivision	188 858	27.43	14.29	0.5831**
40. A. vinelandiisubdivision78 438 65.24 85.93 0.616^{11} 41. B. pertussissubdivision91 147 67.87 85.13 0.7175^{11} 42. N. gonorhoeaesubdivision140 675 52.56 59.92 0.5965^{11} 43. N. mening itidissubdivision33 391 51.12 56.80 0.5768^{11} 44. N. mening itidissubdivision $589 048$ 53.06 61.39 0.5735^{11} 45. N. mening itidissubdivision $583 889$ 53.32 62.08 0.5660^{11} 46. A. nam efacienssubdivision $294 124$ 56.72 64.20 0.7771^{11} 47. A. ham efacienssubdivision $1697 312$ 59.77 71.48 0.7658^{11} 48. A. am efaciens str. C58subdivision $1697 312$ 59.74 71.45 0.7581^{11} 49. R. legun inosa nmsubdivision $86 387$ 60.13 72.71 0.7650^{11} 50. R. hizogenessubdivision $1697 312$ 59.74 71.45 0.5871^{11} 51. R. hizohim sp. NGR234subdivision $143 497$ 68.18 88.07 0.6839^{11} 53. R. sphaenidessubdivision $312 081$ 30.63 18.43 0.5221^{11} 54. R. praw azekiisubdivision $312 081$ 30.63 18.43 0.5221^{11} 55. R. conoriisubdivision $312 081$ 30.63 18.43 0.5221^{11} 56. P. denitrificanssubdivision $322 339$ 62.30 75.90 0.7	38. P. aeruginosa		subdivision	2 313 442	66.44	86.27	0.6446**
41. B. pertussis subdivision 91 147 67. 87 85. 13 0.7175 ^{**} 42. N. gonorrhoeae subdivision 140 675 52. 56 59.92 0.5963 ^{**} 43. N. mening itidis subdivision 303 391 51. 12 56. 80 0.5768 ^{**} 44. N. mening itidis subdivision 589 048 53.06 61.39 0.5735 ^{**} 45. N. mening itidis subdivision 581 889 53.32 62.08 0.5660 ^{**} 46. A. Iam efaciens subdivision 1 668 270 59.77 71.48 0.7658 ^{**} 7U.W ashingon) subdivision 1 697 312 59.74 71.45 0.7881 ^{**} 49. R. legum inosa nm subdivision 1 697 312 59.74 71.45 0.7881 ^{**} 51. Rhizogenes subdivision 163 774 66.38 83.81 0.587 ^{**} 52. R. capsulatus subdivision 113 774 66.38 83.81 0.588 ^{**} 53. R. sphaeroides subdivision 131 2081 30.63 18.43 0.528 ^{**} 54. R. prow azekii subdivision 132 081 30.63 18.43 0	39. P. putida		subdivision	345 718	60.25	73.64	0.6373**
42. N. gonorrhoeae subdivision 140 675 52.56 59.92 0.5963'' 43. N. m eningitidits subdivision 303 391 51.12 56.80 0.5768'' 44. N. m eningitidits subdivision 589 048 53.06 61.39 0.5735'' 45. N. m eningitidits subdivision 589 048 53.22 62.08 0.5660'' 45. N. m enigitidits 294 124 56.72 64.20 0.7771'' 47. A. num efaciens subdivision 1 668 270 59.77 71.48 0.7658'' (U.W ashington) subdivision 1 697 312 59.74 71.45 0.7581'' 48. A. num efaciens str. C58 (Cereon) subdivision 86 387 60.13 72.71 0.7650'' 50. R. rhizogenes subdivision 129 817 58.26 82.40 0.7993'' 51. Rhizohim sp. NGR234 subdivision 129 817 68.18 88.81 0.581'' 53. R. sphaeroides subdivision 312 081 30.63 18.43 0.522'' 53. R. sphaeroides subdivision 322 082 23.58 0.5168'' 54	40. A. vineland ii		subdivision	78 438	65.24	85.93	0.6167**
43. N. m en ing itid issubdivision303 39151.1256.800.5768''44. N. m en ing itid is MC58subdivision589 04853.0661.390.5735''45. N. m en ing itid is Z2491subdivision583 88953.3262.080.5660''46. A. tom efacienssubdivision294 12456.7264.200.7771''47. A. tom efaciens str. C58subdivision1697 31259.7771.480.7558''(U.W ashington)subdivision1697 31259.7471.450.7581''48. A. tom efaciens str. C58 (Cereon)subdivision86 38760.1372.710.7650''50. R. rhizogenessubdivision96 35157.6066.180.8180''51. Rhizohim sp. NCR234subdivision163 77466.3883.810.5871''53. R. sphaeroidessubdivision143 49768.1888.070.6339''54. R. prow azekiisubdivision347 82632.9223.580.5168''55. R. conoriisubdivision137 60269.2489.050.6032''56. P. denitrificanssubdivision137 60269.2489.050.6032''59. H. pylorisubdivision137 60932.8822.930.3711''64. S. aureusBacteriaLow G+C2 783 90844.3144.600.3171'64. S. aureus subsp. aureusNu50EacteriaLow G+C2 783 90844.3144.600.3171'64. S. aureus subsp. aureusNu50Eacteria	41. B. pertussis		subdivision	91 147	67.87	85.13	0.7175**
44. N. m en ing itidisMC58 subdivision 589 048 53. 06 61. 39 0. 5735 ^{**} 45. N. m en ing itidis Z2491 subdivision 583 889 53. 32 62. 08 0. 5660 ^{**} 46. A. tam efaciens subdivision 294 124 56. 72 64. 20 0. 7771 ^{**} 47. A. tam efaciens str. C58 subdivision 1 668 270 59. 77 71. 48 0. 7658 ^{**} (U.W ashingon) subdivision 1 679 312 59. 74 71. 45 0. 7658 ^{**} 48. A. tam efaciens str. C58 (Cereon) subdivision 86 387 60.13 72. 71 0. 7650 ^{**} 50. R. rhizogenes subdivision 129 817 58. 92 68. 24 0. 7993 ^{**} 52. R. capsulatus subdivision 143 497 68. 88 88. 80.7 0. 6839 ^{**} 54. R. prow azekii subdivision 312 081 30. 63 18. 43 0. 5322 ^{**} 55. R. conorii subdivision 232 339 62. 30 75. 90 0. 7422 ^{**} 55. R. conorii subdivision 232 339 62. 30 75. 90 0. 7422 ^{**} 56. P. denitrificans subdivision	42. N. gonorrhoeae		subdivision	140 675	52.56	59.92	0. 5963 * *
45. N. meningitidis Z2491 subdivision 583 889 53. 32 62. 08 0. 5660 ⁺⁺ 46. A. tam efaciens subdivision 294 124 56. 72 64. 20 0. 7771 ⁺⁺ 47. A. tam efaciens str. C58 subdivision 1 668 270 59. 77 71. 48 0. 7658 ⁺⁺ 48. A. nam efaciens str. C58 (Cereon) subdivision 1 697 312 59. 74 71. 45 0. 7581 ⁺⁺ 49. R. legum inosanam subdivision 96 351 57. 60 66. 18 0. 8180 ⁺⁺ 50. R. rhizogenes subdivision 1697 312 59. 74 71. 45 0. 7650 ⁺⁺ 51. Rhizohim sp. NGR234 subdivision 163 774 66. 38 83.81 0. 5871 ⁺⁺ 53. R. sphaeroides subdivision 143 497 68. 18 88.07 0. 6839 ⁺⁺ 54. R. prowazekii subdivision 312 081 30. 63 18. 43 0. 5221 ⁺⁺ 55. R. conorii subdivision 137 602 69. 24 89. 05 0. 6032 ⁺⁺ 55. P. denitrificans subdivision 242 166 31 86.91 0. 5321 ⁺⁺ 59. H. pylori subdivision 137 602	43. $N \cdot m en ing it id is$		subdivision	303 391	51.12	56.80	0. 5768 ^{**}
46. A. um efaciens subdivision 294 124 56. 72 64. 20 0. 7771'' 47. A. tum efaciens str. C58 subdivision 1 668 270 59. 77 71. 48 0. 7658''. 48. A. num efaciens str. C58 (Cereon) subdivision 1 697 312 59. 74 71. 45 0. 7658''. 49. R. legum inosanan subdivision 86 387 60.13 72. 71 0. 7650''. 50. R. rhizogenes subdivision 96 351 57. 60 66. 18 0. 8180''. 51. Rhizobian sp. NGR234 subdivision 163 774 66. 38 83.81 0. 5871''. 53. R. sphaeroides subdivision 131 2081 30.63 18. 43 0. 5281''. 55. R. conorii subdivision 312 081 30.63 18. 43 0. 5322''. 57. B. japonicum subdivision 232 339 62.30 75.90 0. 7422''. 58. M. xanthus subdivision 137 602 69.24 89.05 0.6032''. 59. H. pylori subdivision 137 602 69.24 89.05 0.6047''. 61. H. pylori26695 subdivision 146 741 39.90	44. N.meningitidisMC58		subdivision	589 048	53.06	61.39	0.5735**
47. A. tun efaciens str. C58 subdivision 1 668 270 59.77 71.48 0.7658*. (U.W ashington) subdivision 1 697 312 59.74 71.45 0.7581*. 48. A. tun efaciens str. C58 (Cercon) subdivision 1 697 312 59.74 71.45 0.7658*. 49. R. legum inosanen subdivision 96 351 57.60 66.18 0.8180*. 50. R. hizopienes subdivision 129 817 58.92 68.24 0.7993*. 52. R. capsulatus subdivision 143 497 68.18 88.07 0.683*. 53. R. sphaeroides subdivision 312 081 30.63 18.43 0.5281*. 55. R. conorii subdivision 344 926 32.29 23.58 0.5168* 55. R. conorii subdivision 137 602 69.24 89.05 0.6032* 55. R. conorii subdivision 137 602 69.24 89.05 0.6032* 56. P. denitrificans subdivision 137 602 69.24 89.05 0.6032* 59. H. pylori subdivision 1498 249 39.56 41.95 0.6047*	45. <i>N</i> . <i>m</i> en ing itid is Z2491		subdivision	583 889	53.32	62.08	0.5660**
u.W ashington) subdivision 1 668 270 59. 77 71. 48 0. 7658 48. A. um efaciens str. C58 (Cereon) subdivision 1 697 312 59. 74 71. 45 0. 7658 49. R. legum inosa nun subdivision 86 387 60. 13 72. 71 0. 7650 ⁺⁺ 50. R. rhizogenes subdivision 96 351 57. 60 66. 18 0. 8180 ⁺⁺ 51. Rhizobium sp. NGR234 subdivision 163 774 66. 38 83. 81 0. 5871 ⁺⁺ 52. R. capsulatus subdivision 163 774 66. 38 83. 81 0. 5871 ⁺⁺ 53. R. sphaeroides subdivision 312 081 30. 63 18. 43 0. 5281 ⁺⁺ 54. R. prowazekii subdivision 347 826 32. 92 23. 58 0. 5168 ⁺⁺ 56. P. denitrificans subdivision 574 21 66. 31 86. 91 0. 5322 ⁺⁺ 57. B. japonicum subdivision 137 602 69. 24 89. 05 0. 6032 ⁺⁺ 58. M. xanthus subdivision 265 678 39. 60 39. 80 0. 5601 ⁺⁺⁺ 60. H. pylori 26695 subdivision 495 471 39. 9	46. A. tum efaciens		subdivision	294 124	56.72	64.20	0.7771**
(U.W ashington) $48. A. tam efaciens str. C58 (Cereon)$ subdivision1 697 31259. 7471. 45 0.7581^{**} $49. R. legum inosanam$ subdivision86 38760. 1372. 71 0.7650^{**} $50. R. rhizogenes$ subdivision96 35157. 6066. 18 0.8180^{**} $51. Rhizobiam sp. NGR234$ subdivision163 77466. 3883. 81 0.5871^{**} $52. R. capsulatus$ subdivision163 77466. 3883. 81 0.5871^{**} $53. R. sphaeroides$ subdivision143 49768. 1888. 07 0.6839^{**} $54. R. prow azekii$ subdivision312 08130. 6318. 43 0.5281^{**} $55. R. conorii$ subdivision347 82632. 9223. 58 0.5168^{**} $56. P. den itrificanssubdivision57 42166. 3186. 910.5322^{**}57. B. japonicumsubdivision137 60269. 2489. 050.6032^{**}59. H. py lorisubdivision137 60269. 2489. 050.6047^{**}61. H. py lori 309subdivision495 47139.9042. 660.5967^{**}62. B. subtilisBacteriaLow G+C2 783 90844.3144.600.3171^{*}63. S. aureus subsp. aureus Nu50811 14833.5422. 690.3435^{*}65. S. aureus subsp. aureus N315790 21533. 5122. 530.3402^{**}65. S. aureus subsp. aureus N315790 21533. 5122. 530.3402^{*$	47. A. tum efaciens str. C58			1 669 270	50 77	71 40	0.7650**
49. $R.$ legum inosanomsubdivision86 38760.1372. 710.7650''.50. $R.$ rhizogenessubdivision96 35157.6066.180.8180''.51. R hizobian sp. NGR234subdivision129 81758.9268.240.7993''.52. $R.$ capsulatussubdivision163 77466.3883.810.5871''.53. $R.$ sphaeroidessubdivision143 49768.1888.070.6839''.54. $R.$ provazekiisubdivision312 08130.6318.430.5281''.55. $R.$ conoriisubdivision347 82632.9223.580.5168''.56. $P.$ denitrificanssubdivision57 42166.3186.910.5322''.57. $B.$ japonicumsubdivision232 33962.3075.900.7422''.58. $M.$ xanthussubdivision137 60269.2489.050.6032''.59. $H.$ pylorisubdivision265 67839.6039.800.5601''.60. $H.$ pylori26695subdivision492 4939.5641.950.6047'.61. $H.$ pylori199subdivision495 47139.9042.660.5967''.62. $B.$ subtilisBacteriaLow G+C2 783 90844.3144.600.3171'64. $S.$ aureussubsp. aureusNu50811 14833.5422.690.3435'65. $S.$ naureus subsp. aureusN315790 21533.5122.530.3402'66. $S.$ pneum oniae370 86839.1633.510.2967'67. $L.$ lactis <t< td=""><td>(U.Washington)</td><td></td><td>subdivision</td><td>1 008 270</td><td>59.77</td><td>/1.48</td><td>0. /658</td></t<>	(U.Washington)		subdivision	1 008 270	59.77	/1.48	0. /658
50. R. rhizogenes subdivision 96 351 57. 60 66. 18 0.8180 ^{**} 51. Rhizobium sp. NGR234 subdivision 129 817 58. 92 68.24 0.7993 ^{**} 52. R. capsulatus subdivision 163 774 66.38 83.81 0.5871 ^{**} 53. R. sphaeroides subdivision 143 497 68.18 88.07 0.6839 ^{**} 54. R. prow azekii subdivision 312 081 30.63 18.43 0.5281 ^{**} 55. R. conorii subdivision 347 826 32.92 23.58 0.5168 ^{**} 56. P. denitrificans subdivision 232 339 62.30 75.90 0.7422 ^{**} 58. M. xanthus subdivision 137 602 69.24 89.05 0.6032 ^{**} 59. H. pylori subdivision 1265 678 39.60 39.80 0.5601 ^{**} 60. H. pylori 26695 subdivision 495 471 39.90 42.66 0.5967 ^{**} 61. H. pylori J99 subdivision 495 471 39.90 42.66 0.5967 ^{**} 62. S. aureus Subsp. aureusMu50 811 148 33.51 22.69 0.3435 [*] </td <td>48. A. tum efaciens str. C58 (Ce</td> <td>reon)</td> <td>subdivision</td> <td>1 697 312</td> <td>59.74</td> <td>71.45</td> <td>0.7581**</td>	48. A. tum efaciens str. C58 (Ce	reon)	subdivision	1 697 312	59.74	71.45	0.7581**
51.Rhizobium sp. NGR234subdivision $129\ 817$ $58.\ 92$ $68.\ 24$ $0.\ 7993^{**}$ 52. $R.\ capsulatus$ subdivision $163\ 774$ $66.\ 38$ $83.\ 81$ $0.\ 5871^{**}$ 53. $R.\ sphaeroides$ subdivision $143\ 497$ $68.\ 18$ $88.\ 07$ $0.\ 6839^{**}$ 54. $R.\ prow\ azekii$ subdivision $312\ 081$ $30.\ 63$ $18.\ 43$ $0.\ 5281^{**}$ 55. $R.\ conorii$ subdivision $347\ 826$ $32.\ 92$ $23.\ 58$ $0.\ 5168^{**}$ 56. $P.\ denitrificans$ subdivision $57\ 421$ $66.\ 31$ $86.\ 91$ $0.\ 5322^{**}$ 57. $B.\ japonicum$ subdivision $232\ 339$ $62.\ 30$ $75.\ 90$ $0.\ 7422^{**}$ 58. $M.\ xanthus$ subdivision $137\ 602$ $69.\ 24$ $89.\ 05$ $0.\ 6032^{**}$ 59. $H.\ pylori$ subdivision $265\ 678$ $39.\ 60$ $39.\ 80$ $0.\ 5601^{**}$ 60. $H.\ pylori$ subdivision $498\ 249$ $39.\ 56$ $41.\ 95$ $0.\ 6047^{**}$ 61. $H.\ pylori$ subdivision $495\ 471$ $39.\ 90$ $42.\ 66$ $0.\ 5967^{**}$ 62. $B.\ subtilis$ BacteriaLow $G+C$ $2\ 783\ 908$ $44.\ 31$ $44.\ 60$ $0.\ 3171^{**}$ 64. $S.\ aureus\ subsp.\ aureus\ N150$ 811148 $33.\ 54$ $22.\ 69$ $0.\ 345^{**}$ 65. $S.\ aureus\ subsp.\ aureus\ N315$ $790\ 215$ $33.\ 51$ $22.\ 53$ $0.\ 3402^{**}$ 66	49. R. legum inosa rum		subdivision	86 387	60.13	72.71	0.7650**
52. R. capsulatussubdivision163 77466. 3883. 810. 5871''.53. R. sphaeroidessubdivision143 49768. 1888. 070. 6839''.54. R. prow azekiisubdivision312 08130. 6318. 430. 5281''.55. R. conoriisubdivision347 82632. 9223. 580. 5168''.56. P. den itrificanssubdivision57 42166. 3186. 910. 5322''.57. B. japonicumsubdivision232 33962. 3075. 900. 7422''.58. M. xanthussubdivision137 60269. 2489. 050. 6032''.59. H. pylorisubdivision265 67839. 6039. 800. 5601''.60. H. pylori26695subdivision498 24939. 5641. 950. 6047''.61. H. pyloriJ99subdivision495 47139.9042. 660. 5967''.62. B. subtilisBacteriaLow G+C2 783 90844. 3144. 600. 3171''.64. S. aureussubsp. aureusMu50811 14833. 5422. 690. 3435'65. S. aureussubsp. aureusN315790 21533. 5122. 530. 3402''66. S. pneum oniae370 86839. 1633. 510. 2967''67. L. lactis233 61735.4926. 410. 3739''68. E. faecalis145 56237. 8131. 650. 2875'69. M. genitaliumMycoplasma181 43531. 7423. 240. 6631''	50. R. rhizogenes		subdivision	96 351	57.60	66.18	0.8180**
53. R. sphaeroidessubdivision143 49768. 1888. 07 0.6839^{++} 54. R. prow azek iisubdivision312 08130. 6318. 43 0.5281^{++} 55. R. conoriisubdivision347 82632. 9223. 58 0.5168^{++} 56. P. denitrificanssubdivision57 42166. 3186. 91 0.5322^{++} 57. B. japonicumsubdivision232 33962. 3075. 90 0.7422^{++} 58. M. xanthussubdivision137 60269. 2489. 05 0.6032^{++} 59. H. pylorisubdivision265 67839. 6039. 80 0.5601^{++} 60. H. pylori 26695subdivision498 24939. 5641. 95 0.6047^{++} 61. H. pylori J99subdivision495 47139. 9042. 66 0.5967^{++} 62. B. subtilisBacteriaLow G+C2 783 90844. 3144. 60 0.3171^{++} 63. S. aureusGram-positive431 60932. 8822. 93 0.3711^{++} 64. S. aureus subsp. aureusN315790 21533. 5122. 53 0.3402^{+} 65. S. pneum oniae370 86839. 1633. 51 0.2967^{+} 67. L. lactis233 61735. 4926. 41 0.3739^{++} 68. E. faecalis145 56237. 8131. 65 0.2875^{+} 69. M. genitaliumMycop lasma181 43531. 7423. 24 0.6631^{++}	51. Rhizobium sp. NGR234		subdivision	129 817	58.92	68.24	0. 7993**
54. $R. prow azekii$ subdivision312 08130. 6318. 43 0.5281° 55. $R. conorii$ subdivision $347 826$ $32. 92$ $23. 58$ 0.5168° 56. $P. den itrificans$ subdivision $57 421$ $66. 31$ $86. 91$ 0.5322° 57. $B. japonicum$ subdivision $232 339$ $62. 30$ $75. 90$ 0.7422° 58. $M. xanthus$ subdivision $137 602$ $69. 24$ $89. 05$ 0.6032° 59. $H. py lori$ subdivision $265 678$ $39. 60$ $39. 80$ 0.5601° 60. $H. py lori 26695$ subdivision $498 249$ $39. 56$ $41. 95$ 0.6047° 61. $H. py lori J99$ subdivision $495 471$ $39. 90$ $42. 66$ 0.5967° 62. $B. subtilis$ BacteriaLow $G + C$ $2783 908$ $44. 31$ $44. 60$ 0.3171° 63. $S. aureus$ Gram-positive $431 609$ $32. 88$ $22. 93$ 0.3711° 64. $S. aureus subsp. aureus Mu50$ 811 148 $33. 51$ $22. 53$ 0.3402° 65. $S. pneum oniae$ $370 868$ $39. 16$ $33. 51$ 0.2967° 67. $L. lactis$ $233 617$ 35.49 26.41 0.3739° 68. $E. faecalis$ $145 562$ $37. 81$ $31. 65$ 0.2875° 69. $M. genitalium$ Mycop lasna181 435 $31. 74$ $23. 24$ 0.6631°	52. R. capsulatus		subdivision	163 774	66.38	83.81	0.5871**
55. R. conoriisubdivision $347\ 826$ $32.\ 92$ $23.\ 58$ $0.\ 5168^{**}$ 56. P. denitrificanssubdivision $57\ 421$ $66.\ 31$ $86.\ 91$ $0.\ 5322^{**}$ 57. B. japonicumsubdivision $232\ 339$ $62.\ 30$ $75.\ 90$ $0.\ 7422^{**}$ 58. M. xanthussubdivision $137\ 602$ $69.\ 24$ $89.\ 05$ $0.\ 6032^{**}$ 59. H. pylorisubdivision $265\ 678$ $39.\ 60$ $39.\ 80$ $0.\ 5601^{**}$ 60. H. pylori 26695subdivision $498\ 249$ $39.\ 56$ $41.\ 95$ $0.\ 6047^{**}$ 61. H. pylori J99subdivision $495\ 471$ $39.\ 90$ $42.\ 66$ $0.\ 5967^{**}$ 62. B. subtilisBacteriaLow G+C $2\ 783\ 908$ $44.\ 31$ $44.\ 60$ $0.\ 3171^{*}$ 63. S. aureusGram-positive $431\ 609$ $32.\ 88$ $22.\ 93$ $0.\ 3711^{**}$ 64. S. aureus subsp. aureusNu50811148 $33.\ 54$ $22.\ 69$ $0.\ 3402^{**}$ 65. S. naureus subsp. aureusN315790\ 215 $33.\ 51$ $0.\ 2967^{**}$ 66. S. pneum oniae370\ 868 $39.\ 16$ $33.\ 51$ $0.\ 2967^{**}$ 67. L. lactis145\ 562 $37.\ 81$ $31.\ 65$ $0.\ 2875^{**}$ 69. M. genitaliumMycop lasna181435 $31.\ 74$ $23.\ 24$ $0.\ 6631^{**}$	53. R. sphaeroides		subdivision	143 497	68.18	88.07	0.6839**
56. P. den itrificanssubdivision57 42166. 3186. 91 0.5322^{**} 57. B. japonicumsubdivision232 33962. 3075. 90 0.7422^{**} 58. M. xan thussubdivision137 60269. 2489.05 0.6032^{**} 59. H. pylorisubdivision265 67839. 6039. 80 0.5601^{**} 60. H. pylori 26695subdivision498 24939. 5641.95 0.6047^{**} 61. H. pylori J99subdivision495 47139. 9042. 66 0.5967^{**} 62. B. subtilisBacteriaLow G+C2 783 90844. 3144. 60 0.3171^{*} 63. S. aureusGram-positive431 60932. 8822. 93 0.3711^{**} 64. S. aureus subsp. aureusMu50811 14833. 5422. 69 0.3435^{*} 65. S. naureus subsp. aureusN315790 21533. 5122. 53 0.3402^{**} 66. S. pneum oniae370 86839.1633. 51 0.2967^{**} 67. L. lactis145 56237. 8131.65 0.2875^{**} 69. M. genitaliumMycop lasma181 43531.7423. 24 0.6631^{**}	54. R. prowazekii		subdivision	312 081	30.63	18.43	0.5281**
$57. B. japon icum$ subdivision $232 339$ $62. 30$ $75. 90$ $0. 7422^{**}$ $58. M. xan thus$ subdivision $137 602$ $69. 24$ $89. 05$ $0. 6032^{**}$ $59. H. py lori$ subdivision $265 678$ $39. 60$ $39. 80$ $0. 5601^{**}$ $60. H. py lori 26695$ subdivision $498 249$ $39. 56$ $41. 95$ $0. 6047^{**}$ $61. H. py lori J99$ subdivision $495 471$ $39. 90$ $42. 66$ $0. 5967^{**}$ $62. B. subtilis$ BacteriaLow $G+C$ $2 783 908$ $44. 31$ $44. 60$ $0. 3171^{*}$ $63. S. au reus$ Gram -positive $431 609$ $32. 88$ $22. 93$ $0. 3711^{**}$ $64. S. au reus subsp. au reus Mu50811 14833. 5422. 690. 3435^{*}65. S. pneum on iae370 86839. 1633. 510. 2967^{*}68. E. faecalis145 56237. 8131. 650. 2875^{*}69. M. genitaliumMycop lasma181 43531. 7423. 240. 6631^{**}$	55. R. conorii		subdivision	347 826	32.92	23.58	0.5168**
$58. M. xan thus$ subdivision $137\ 602$ $69.\ 24$ $89.\ 05$ $0.\ 6032^{**}$ $59. H. pylori$ subdivision $265\ 678$ $39.\ 60$ $39.\ 80$ $0.\ 5601^{**}$ $60. H. pylori\ 26695$ subdivision $498\ 249$ $39.\ 56$ $41.\ 95$ $0.\ 6047^{**}$ $61. H. pylori\ J99$ subdivision $495\ 471$ $39.\ 90$ $42.\ 66$ $0.\ 5967^{**}$ $62. B.\ subtilis$ BacteriaLow G+C $2\ 783\ 908$ $44.\ 31$ $44.\ 60$ $0.\ 3171^{*}$ $63.\ S.\ aureus$ Gram-positive $431\ 609$ $32.\ 88$ $22.\ 93$ $0.\ 3711^{**}$ $64.\ S.\ aureus\ subsp.\ aureus\ Mu50$ 811\ 148 $33.\ 54$ $22.\ 69$ $0.\ 3435^{*}$ $65.\ S.\ aureus\ subsp.\ aureus\ N315$ 790\ 215 $33.\ 51$ $22.\ 53$ $0.\ 3402^{*}$ $66.\ S.\ pneum\ oniae$ 370\ 868 $39.\ 16$ $33.\ 51$ $0.\ 2967^{*}$ $68.\ E.\ faecalis$ 145\ 562 $37.\ 81$ $31.\ 65$ $0.\ 2875^{*}$ $69.\ M.\ genitalium$ Mycop lasma181\ 435 $31.\ 74$ $23.\ 24$ $0.\ 6631^{**}$	56. P. denitrificans		subdivision	57 421	66.31	86.91	0.5322**
59. H. pylorisubdivision265 67839. 6039. 800. 5601**60. H. pylori 26695subdivision498 24939. 5641. 950. 6047**61. H. pylori J99subdivision495 47139. 9042. 660. 5967**62. B. subtilisBacteriaLow G+C2 783 90844. 3144. 600. 3171*63. S. aureusGram-positive431 60932. 8822. 930. 3711**64. S. aureus subsp. aureusMu50811 14833. 5422. 690. 3402*65. S. aureus subsp. aureusN315790 21533. 5122. 530. 3402*66. S. pneum oniae370 86839. 1633. 510. 2967*68. E. faecalis145 56237. 8131. 650. 2875*69. M. genitaliumMycop lasma181 43531. 7423. 240. 6631**	57. B. japonicum		subdivision	232 339	62.30	75.90	0.7422**
$60. H. py lori 26695$ subdivision $498 249$ 39.56 41.95 0.6047^{**} $61. H. py lori J99$ subdivision $495 471$ 39.90 42.66 0.5967^{**} $62. B. subtilis$ BacteriaLow G+C $2.783 908$ 44.31 44.60 0.3171^{*} $63. S. au reus$ Gram positive $431 609$ 32.88 22.93 0.3711^{**} $64. S. au reus subsp. au reus Mu50811 14833.5422.690.3435^{*}65. S. au reus subsp. au reus N315790 21533.5122.530.3402^{*}66. S. pneum oniae370 86839.1633.510.2967^{*}68. E. faecalis145 56237.8131.650.2875^{*}69. M. genitaliumMycop lasma181 43531.7423.240.6631^{**}$	58. M. xanthus		subdivision	137 602	69.24	89.05	0.6032**
$60. H. py lori 26695$ subdivision $498 249$ 39.56 41.95 0.6047^{**} $61. H. py lori J99$ subdivision $495 471$ 39.90 42.66 0.5967^{**} $62. B. subtilis$ BacteriaLow G+C $2.783 908$ 44.31 44.60 0.3171^{*} $63. S. au reus$ Gram positive $431 609$ 32.88 22.93 0.3711^{**} $64. S. au reus subsp. au reus Mu50811 14833.5422.690.3435^{*}65. S. au reus subsp. au reus N315790 21533.5122.530.3402^{*}66. S. pneum oniae370 86839.1633.510.2967^{*}68. E. faecalis145 56237.8131.650.2875^{*}69. M. genitaliumMycop lasna181 43531.7423.240.6631^{**}$	59. H. pylori		subdivision	265 678	39.60	39.80	0.5601**
62. B. subtilisBacteriaLow G+C2 783 90844. 3144. 60 0.3171° 63. S. aureusGram-positive431 60932. 8822. 93 0.3711° 64. S. aureus subsp. aureusMu50811 14833. 5422. 69 0.3435° 65. S. aureus subsp. aureusN315790 21533. 5122. 53 0.3402° 66. S. pneum oniae370 86839. 1633. 51 0.2967° 67. L. lactis233 61735. 4926. 41 $0.3739^{\circ*}$ 68. E. faecalis145 56237. 8131. 65 0.2875° 69. M. genitaliumMycop lasna181 43531. 7423. 24 $0.6631^{\circ*}$	60. H. pylori 26695		subdivision	498 249	39.56	41.95	
63. S. aureusGram-positive431 60932. 8822. 930. 3711**64. S. aureus subsp. aureusMu50811 14833. 5422. 690. 3435*65. S. aureus subsp. aureusN315790 21533. 5122. 530. 3402*66. S. pneum oniae370 86839. 1633. 510. 2967*67. L. lactis233 61735. 4926. 410. 3739**68. E. faecalis145 56237. 8131. 650. 2875*69. M. genitaliumMycop lasna181 43531. 7423. 240. 6631**	61. H. pylori J99		subdivision	495 471	39.90	42.66	0. 5967 * *
64. S. aureus subsp. aureusMu50811 14833. 5422. 690. 3435*65. S. aureus subsp. aureusN315790 21533. 5122. 530. 3402*66. S. pneum oniae370 86839. 1633. 510. 2967*67. L. lactis233 61735. 4926. 410. 3739**68. E. faecalis145 56237. 8131. 650. 2875*69. M. genitaliumMycop lasma181 43531. 7423. 240. 6631**	62. B. subtilis	Bacteria	Low $G + C$	2 783 908	44.31	44.60	
65. S. aureus subsp. aureus N315790 21533. 5122. 530. 3402*66. S. pneum oniae370 86839. 1633. 510. 2967*67. L. lactis233 61735. 4926. 410. 3739**68. E. faecalis145 56237. 8131. 650. 2875*69. M. genitaliumMycop lasma181 43531. 7423. 240. 6631**	63. S. aureus		Gram-positive	431 609	32.88	22.93	0.3711**
66. S. pneum oniae370 86839. 1633. 510. 2967*67. L. lactis233 61735. 4926. 410. 3739**68. E. faecalis145 56237. 8131. 650. 2875*69. M. genitaliumMycoplasna181 43531. 7423. 240. 6631**	64. S. aureus subsp. aureus Mu50)		811 148	33.54	22.69	0.3435*
66. S. pneum oniae370 86839. 1633. 510. 2967*67. L. lactis233 61735. 4926. 410. 3739**68. E. faecalis145 56237. 8131. 650. 2875*69. M. genitaliumMycoplasna181 43531. 7423. 240. 6631**	65. S. aureus subsp. aureus N315	j			33.51	22.53	
67. L. lactis233 61735. 4926. 410. 3739**68. E. faecalis145 56237. 8131. 650. 2875*69. M. genitaliumMycoplasma181 43531. 7423. 240. 6631**	-				39.16	33.51	
68. E. faecalis145 56237. 8131. 650. 2875*69. M. genitaliumMycoplasma181 43531. 7423. 240. 6631**				233 617	35.49	26.41	
69. M. genitalium Mycop lasma 181 435 31.74 23.24 0.6631	68. E. faecalis			145 562			
			M ycop la sm a				
	70. M. pulm on is		M ycop la sm a	302 999	27.14	14.72	0. 5914 ^{**}

GC, percentage of guanine + cytosine; GC3s, frequency of guanine + cytosine at the synonymous third positions of codons; r is correlation coefficient of 58 codons with their complementary codons (three term in al codons and their corresponding complementary codons are not included); *, * * Significance at the 5% and 1% probability levels respectively.

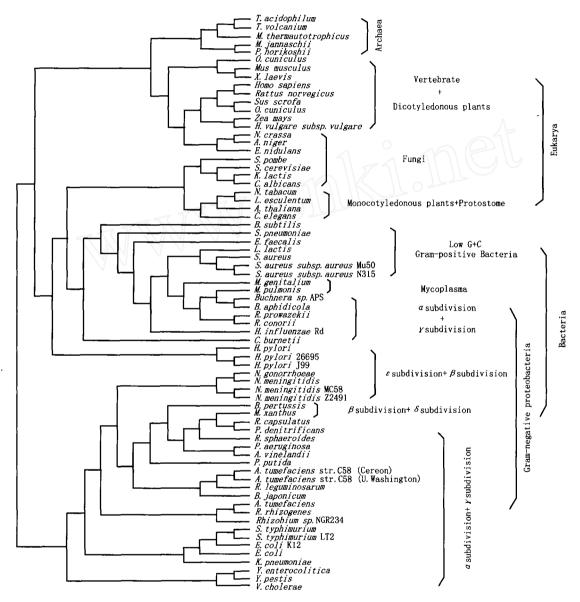
2 Results and discussion

The analysis results of codons usage with their corresponding complimentary ones of genomes from 70 organisms were summarized in Table 1. From Table 1, one could easily see that codons usage has a highly significant or significant positive correlation with their corresponding complementary codons in 70 organisms. The high average correlation coefficients of codons vs complementary codons of genomes from 70 organisms, 0. 5101 with the standard deviation 0. 1389, indicates that the positive usage correlation relationship of codons with their corresponding complementary ones do objectively exist

However, there was also strong heterogeneity of correlation coefficients among 70 organisms (Table 1), For Archaea, the range of correlation coefficient between all codons vs complementary ones usage is from 0.2914 to 0.4027 with a mean of 0.3565 and standard deviation of 0.0468; while in Eukaryota, from 0.2965 to 0.5089 with a mean of 0. 4460 and standard deviation of 0. 0665; and in Bacteria, from 0. 2800 to 0. 8180 with a mean of 0. 5556 and standard deviation of 0. 1466. The mean of correlation coefficient from 45 Bacteria is rather higher than that from A rchaea and Eukaryota It was further found that there exists general descending tendency of their means of correlation coefficient: subdivision (0.6895) > M y coplasm a Bacteria (0.6272) >subdivision subdivision (0.5872) >(0.6060) >subdivision (0.6032) >subdivision (0.5167) > Low (G+C)Gram-positive Bacteria (0.3329) in Bacteria dataset, while in Eukaryota dataset: Fungi (0.4637) > Vertebrates (0.4501) > Plants (0.4161). The results suggest that the correlation coefficient of codons vs complimentary ones usage to some extent not only reflect the evolutionary history of these organisms, but also reveal the co-evolution processing of codons vs complimentary codons in organisms Specially, for the mophiles from Archaea, M. jannaschii, M. theim autotrophicus, P. horikoshii, T. acidophilum and T. volcanium grow at environto over 100 ^[10]. Withstanding higher temperatures, they show that ment of relative high temperature from 60 their bio-molecules composition, especially proteins, must have been changed for normal biological function Therefore, in order to maintain the kind of function stability, these thermophiles should have endured a much stronger selective pressure during their evolution This kind of selection pressure may have intensively twisted the correlation between codons and their complementary codons usage, and result in a lower correlation coefficient Hence, the correlation extent of codons vs their complementary codons usage may be used as a fair index for measuring the degree of selection force during organism evolution

In order to determine the effect of GC content on correlation of all codons with complimentary ones usage, the relation analysis was carried out between G + C content and correlation coefficient of all codons vs complimentary ones usage among the 70 genomes The results show a highly significant positive correlation between GC content and correlation coefficient (r = 0.5230, P < 0.0001) as well as GC3 s content and correlation coefficient (r = 0.5307, P < 0.0001). The higher correlation coefficient is with higher GC content of the complete genome of organism. Therefore GC and GC3s content also reveal strong effects upon the correlation between all codons vs complementary ones usage. Recently it was suggested that compositional constraints are the main factors of deciding the codon usage changes among the genes and organism s^[11,12], and that the combination of translational selection and compositional constraints acts for dictating the codon usage variation among genes^[13,14]. However, among prokaryotes, it appears that the influences of natural selection and mutational biases are different if the genome is skewed towards AT or GC, and the analyses of the completed genomes of Rickettsia prowazekii and Borrelia burgdorferi with a genomic GC level of 29% show that the mutational bias is the dominant factor shaping codon usage, while in Mycobacterium tubercubsis (GC = 65%) translational selection on codon choices has been displayed [11, 15 - 19]. Therefore, the present results suggested that the correlation degree of codons with complementtary ones may provide a useful mark for distinguishing the pattern of codon usage among different organisms For example, Romero, et al ^[13] found that compositional pressure and translational selection determine codon usage in the extremely GC-poor unicellular eukaryote Entamoeba histolytica Compositional pressure and translational

selection may also be detected by correlation analysis of codons with complementary ones based on the composition of both codons and their complimentary ones, for both are virtually the same either in GC or AU content If the usage of codons tends to be selected by the composition of the genome, the usage of the complementary codons should accordingly be selected by the same factor. Such usage relationship of codons and complementary ones under compositional pressure and translational selection could then be reflected by their correlation coefficient



The codon-complementary codon tree was built by the difference values between every single pair usage frequency of codon vs complementary codon, and five frequency combination datasets, namely double low frequency numbers (the frequency of codon and its complementary codon < 10‰), double high frequency numbers (the frequency of codon and its complementary codon < 10‰), single low frequency numbers (one of the frequency of codon and its complementary codon < 10‰), single high frequency numbers (one of the frequency of codon and its complementary codon < 10‰), single high frequency numbers (one of the frequency of codon and its complementary codon < 10‰), single high frequency numbers (one of the frequency of codon and its complementary codon > 40‰), and the others (the frequency of codon and its complementary codon between 10‰ and 40‰) in the genomes of 70 organisms D istance between two species were estimated on the 37-dimensional vectors (32 the difference values between every single pair usage frequency of codon vs complementary codon and 5 frequency combination numbers), with each axis representing a pair W ith the parsed data, a distance matrix was then calculated by using M inkowski distance. The codon-complementary codon tree was subsequently constructed by using the Pairwise D istance program with the Neighbor-Joining method in the Mega2 Software^[23].

Figure 1 The codon-complementary codon tree

Theoretically, any genetic codons with their corresponding complementary ones in usage should be complete-

ly positive correlated (namely their correlation coefficients were close to 1). In fact, it has proved that the first base of many anticodons is usually modified and its pairing with the third base of corresponding codon usually wobbled^[20, 21]. In these cases, the equality of the 'codon-anticodon 'vs 'complementary codon-anticodon ' in pairing energy can not be preserved, which results in a biased correlation relationship between them. That is why the calculated correlation coefficients of codons with complementary ones in usage are always much less than 1. The wobbling and the modification in anticodons may directly result in the result of lower correlation coefficients as indicated in our results, and the stronger the wobbling as well as the modification are, the smaller the correlation coefficients will become. This implies correlation coefficients of codons with complementary ones with complementary ones could also be used as the index of anticodons ' modification and wobbling, and may provide much useful information about all anticodons ' modification and wobbling codons with corresponding codons within this organism.

As it is true for the pair usage information of codons vs complementary codons in an organism may imply the evolutionary history of the organism, the neighbouring species in evolution may have similar pair usage patterns of codons vs complementary codons. In order to test the validity of our hypothesis, we built a codon-complementary codon tree based on 32 single pair usage of codon vs complementary codon (Fig 1). As expected, the five Ar-chae bacteria, twenty Eukarya and forty five Bacteria are grouped together respectively, and *M. pulmonis* is nearly grouped with *M. genitalium*, besides the placement of dicotyledonous plants and monocotyledonous plants as well as Fungi was slightly dispatched, the codon-complementary codon tree exhibited very high fidelity to the phylogenetic lineage of the organism s as shown in^[22] Figure 1 and Table 1. This suggests that the usage relation-ship of codons vs complementary codon in organism may strongly reflect their co-evolution history.

In summary, the significantly positive correlation of all codons with their complementary ones in usage among 70 organisms not only proves the validity of our assumption that when a codon is biased in usage, its complementary codon is also biased, but also demonstrates the rationality of Henri Grosjean s ' optimal combination of codon-anticodon complex $\frac{1}{2}$ and lkemura s ' translation efficiency $\frac{1}{2}$, hypotheses in a new point of view. This manifests the usefulness of correlation analysis between codon and complementary codon in studying pattern of codon usage bias, especially for elucidating molecular evolutionary mechanisms

References

- [1] Grosjean H, Sankoff D and Fiers W. Bacteriophage MS2 RNA: A correlation between the stability of the codon: anticodon interaction and the choice of code words [J]. J Mol Evol, 1978, 12: 113 ~ 119
- [2] Grosjean H and Fiers W. Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes [J]. Gene, 1982, 18: 199 ~ 209
- [3] Ikemura T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system [J]. J Mol B iol, 1981, 151: 389 ~ 409
- [4] Ikemura T. Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in proteins genes Differences in synonymous codon choice patterns of yeast and *Escherichia coli* with reference to the abundance of isoaccepting transfer RNAs[J]. J Mol B iol, 1982, 158: 573 ~ 587
- [5] kemura T. Codon usage and tRNA content in unicellular and multicellular organisms [J]. Mol B iol Evol, 1985, 2: 13 ~ 34
- [6] Ikemura T Correlation between codon usage and tRNA content in microorganisms [A]. In: Transfer RNA in Protein Synthesis
 [M]. dDL Hatfield, B J Lee and R M Pirtle eds Boca Raton, FL, CRC Press, 1992. 87 ~ 111
- [7] W right F. The effective number of codons used by a gene[J]. Gene, 1990, 87: 23 ~ 29
- [8] Karlin S and Mr ázek J. What drives codon choices in human genes? [J]. J Mol Biol, 1996, 262: 459 ~ 472
- [9] Kim C H, Oh Y and Lee T H. Codon op timization for high-level expression of human erythropoietin (EPO) in mammalian cells[J]. Gene, 1997, 199: 293 ~ 301
- [10] de Farias S and Bonato M. Preferred codons and am ino acid couples in hyperthermophiles [J]. Genome Biology, 2002, 3: 1 ~ 18, preprint
- [11] Andersson S, Zomorodipour Andersson J, Sicheritz-Ponten T, et al The genome sequence of Rickettsia prow azekii and the ori-

314

gin of mitochondria [J]. Nature, 1998, 396: 133 ~ 140

- [12] Musto H, Romero H, Zavala A, et al Synonymous codon choices in the extremely GC-poor genome of plasmodium falciparum: compositional constraints and translational selection [J]. J Mol Evol, 1999, 49: 27 ~ 35
- [13] Romero H, Zavala A and Musto H. Compositional pressure and translational selection determine codon usage in the extremely GC-poor unicellular eukaryote Entamoeba histolytica [J]. Gene, 2000, 242: 307 ~ 311
- [14] Ghosh T C, Gupta S K and M ajum dar S Studies on codon usage in Entam oeba histolytica [J]. Int J Parasitol, 2000, 30: 725 ~ 722
- [15] Naya H, Romero H, Carels N, et al Translational selection shapes codon usage in the GC-rich genome of Chlam ydom on as reinhardtii [J]. FEBS Lett, 2000, 501: 127 ~ 130
- [16] Andersson S and Sham P. Codon usage and base composition in Rickettsia prowazekii [J]. J Mol Evol, 1996, 42: 525 ~ 536
- [17] Lafay B, Lloyd A, McLean M, et al Protostome composition and codon usage in spirochaetes: species-specific and DNA strand-specific mutational biases [J]. Ncleic Acids Res, 1999, 27: 1642 ~ 1649
- [18] Mc Inemey J O. Replicational and transcriptional selection on codon usage in *Borrelia burgdorferi* [J]. Proc Natl Acad Sci USA, 1998, 95: 10698 ~ 10703
- [19] De Miranda A, Alvarez-Valin F, Jabbari K, et al Gene expression, am ino acid conservation and hydrophobicity are the main factors shaping codon preferences in Mycobacterium tuberculosis and Mycobacterium leprae [J]. J Mol Evol, 2000, 50: 45 ~ 55
- $[\,20\,]$ Crick F H C. The origin of the genetic code[J]. J MolBiol, 1968, 38: 367 ~79
- [21] Yoshiki A, Fumiaki Y, Akira M, et al Codon recognition patterns as deduced from sequences of the complete set of transfer RNA species in M ycoplasm a capricolum [J]. Resemblance to mitochondria J Mol Biol, 1989, 209: 37 ~ 54
- [22] Kirlin S, Campbell A M and M r ázek J. Comparative DNA analysis across diverse genomes [J]. Annu Rev Genet, 1998, 32: 185 ~ 225
- [23] Kumar S, Tamura K, Jakobsen IB, et al MEGA2: molecular evolutionary genetics analysis software [J]. Bioinformatics, 2001, 17: 1244 ~ 1245

基于密码子与互补密码子使用的进化树的重构

黄惠芳¹,庄永龙²,陈礼明¹,林鲁萍¹,马 飞³,许晓风⁴

- (1. 厦门大学生命科学学院,厦门 361005; 2. 清华大学生物信息研究所,北京 100084;
 - 3. 辽宁师范大学生命科学学院比较基因组与生物信息学实验室,大连 116029;
- 4. 南京师范大学生命科学学院,南京 210097)

摘 要:为了更深入地了解密码子与互补密码子使用相关的内涵,70个单细胞和多细胞生物被更进 一步的分析。结果得到 70种细胞生物具有较高的相关系数平均值 (0.5101,标准差是 0.1389),表明密码 子与其互补密码子使用间确实存在正的使用相关联系,进一步地支持了以前的结果。此外,通过对 70个 单细胞和多细胞生物在基于密码子与互补密码子使用频率差异,以及配对频率的高低信息的分析基础上, 绘制进化树,得到的进化树与传统的进化树非常接近。

关键词:密码子;互补密码子;使用频率;使用偏爱;相关分析;密码子 - 互补密码子树