The constraints between amino acids influence the unequal distribution of codons and protein sequence evolution

Yi Qian[1], Rui Zhang[2], Xinglu Jiang[2], and Guoqiu Wu[1]

[1]Southeast University Zhongda Hospital
[2]Southeast University

May 5, 2020

Abstract

4 nucleotides(A, U, C, G) constitute 64 codons at free combination but 64 codons are unequally assigned to 21 items (20 amino acids plus one stop). About 500 amino acids are known but only 20 ones are selected to make up the proteins. However, the relationships between amino acid and codon and between 20 amino acids have been unclear. In this paper, we studied on the relationships between 20 amino acids in 33 species and found there were three constraints between 20 amino acids, such as the relatively stable mean carbon and hydrogen(C:H) ratios(0.50), similarity interactions between the constituent ratios of amino acids, and the frequency of amino acids according with Poisson distribution under a certain conditions. We demonstrated that the unequal distribution of 64 codons and the choice of amino acids in molecular evolution would be constrained to remain stable C:H ratios. The constituent ratios and frequency of 20 amino acids in a species or a protein are two determinants of protein sequence evolution, so this findings showed the constraints between 20 amino acids played an important role in protein sequence evolution.

Introduction

The rates and sequences of 20 amino acids in proteins have remained a central subject in evolutionary and molecular biology for half a century[1-4]. By far, protein expression level and the functional importance of a protein have been viewed to 2 major determinant[5]. About 500 amino acids are known but only 20 ones are selected to make up the proteins, an important sort of biological polymers[6]. Darwin's theory of evolution and neutral theory, two major rival theories, can be used to explain some laws of species evolution or has been justified by some natural phenomena[7-9]. The rate of amino acid reflects both Darwinian selection for functionally advantageous mutations and selectively neutral evolution operating within the constraints of structure and function[10]. However, relationships and interactions between amino acids in molecular evolution are scarcely reported.

Codons, three nucleotides, locate in transfer RNA (tRNA) molecules to carry amino acids and to read the mRNA at a time. 4 nucleotides(A, U, C, G) constitute 64 codons at free combination. However, 64 codons are unequally assigned to 21 items (20 amino acids plus one stop). The explanations about this phenomenon are unclear, such as frozen accident hypothesis[11], stereochemical hypothesis[12], co-evolution hypothesis[13], ATP-centric hypothesis[14] and so on. Based on these observations, we should found a point to establish relationship between 20 amino acids and their corresponding codons and the carbon and hydrogen(C:H) ratios were proved to be a good choice by repeated simulation calculation in this paper.

The molecular evolutionary clock is that the rate of evolution at the molecular level is approximately constant through time and among species[15]. Biologists can compare protein sequences, such as haemoglobins, cytochrome c and fibrinopeptides from different species of mammals, to infer the dates of major species divergence events in the Tree of Life[16-18]. However, multiple factors were found to influence the varying molecular

evolutionary rates among species, which could lead the clock to be violated, including generation time, population size, basal metabolic rate and so on[15,19,20]. Next-generation sequencing technologies have led to the increased availability of genomic data offering molecular clock dating studies and some effective methods also have been reported, such as relaxed clock models, but still the divergences doesn't disappear[21,22]. To find the relationship between 20 amino acids or approximately constant of molecular evolution under what conditions could be helpful to better application of molecular clock.

To seek the similarities of organisms on DNA and protein level should be a method to research on their relationships. Life on earth probably began 3.5-4 billion years ago[23]. Some similarities have been retained for such long time of evolution and it definitely needs constraint forces[5,24-26]. We collected protein sequences from 33 species of genome date from NCBI(See Table S1) and made some statistical analysis to seek these similarities. In this paper, we reported three similarities of 33 species in molecular evolution, which was defined as the constraints between 20 amino acids in protein evolution.

Results and Discussion

The amino acid gain and loss in protein evolution was reported to be a universal trend via comparing 12 available triplets of complete prokaryotic genomes and not to be driven by any simple trend at the DNA level[27]. Here, we set the relationships between the sum of 20 amino acids of all proteins in a species and the number of their corresponding codons to explore the amino acid gain and loss (see Methods). Under ideal condition, there were a linear relation between the two after random mutation, so we made the rate of their corresponding codons as a reference standard to compare the amino acid gain and loss. In Figure 1a, we found 4 amino acids, such as Met(M, CG%=33.3%), Asp(D, CG%=50%), Glu(E, CG%=50%), and Phe(F, CG%=16.7%) in all 33 species were on the red line(y=0), which showed that they are "gainers" , while 7 amino acids, such as Trp(W, CG%=66.7%), Cys(C, CG%=50%), His(H, CG%=50%), Thr(T, CG%=50%), Pro(P, CG%=83.3%), Ser(S, CG%=50%), and Arg(R, CG%=72.2%) in all 33 species were under the red line(y=0), which showed that they are "loser". The biggest variation is Lys(K, from -0.181 to 1.68, CG%=16.7%) and the smallest variation is Thr(T, from -0.413 to -0.181, CG%=50%). GC% of 4 gainers are [?]50% and GC% of 7 loser are [?]50%, which showed protein evolution was driven by any simple trend at the DNA level and more amino acids encoded by (G+C)-rich codons.

Carbon, hydrogen, nitrogen, oxygen and sulfur composition of 20 amino acids constitute variety of proteins. The carbon and hydrogen(C:H) ratios are from 0.40 (Gly, 2/5)to 0.92(Trp, 11/12) and the mean C:H ratio of 20 amino acids is 0.54, while it is 0.50(theoretical value), if the constituent ratio of 20 amino acids happens to be the rates of their corresponding codons in a protein. Then, we respectively calculated the sum of 20 amino acids of all proteins in each species and calculated the mean C:H ratios of each species. However, regardless of the significant change rates of 20 amino acids in different species(see Figure 1a), the carbon and hydrogen ratios kept on a relatively stable value(0.50 ± 0.002, see Figure 1b). It was miraculous that although loss and gain of amino acids resulted in the significant change of their constituent ratios, C:H ratios of 33 species remained theoretical value(0.50). Here, we suspected carbon and hydrogen ratios was a constraint between 20 amino acids in protein evolution and a determinant of the rate of protein sequence evolution.

To further validate the authenticity of this conclusion, we obtained the bivariate correlations of 20 amino acids in 33 species shown in Figure 2a. We found Trp(W, C:H 0.917, the highest C:H ratio) was negative correlation to Phe(F, C:H 0.818, the second-highest C:H ratio) and Tyr(Y, C:H 0.818, the second-highest C:H ratio) and positive correlation to Gly(G, C:H 0.4, the lowest C:H ratio, r=-0.705, p=0.000), Ala(A , C:H 0.429, r=-0.641, p=0.000) and Arg(R, C:H 0.429, r=-0.843,p=0.000). All amino acids with C:H ratios more than 0.5 were found to be positive correlation with at least one amino acid with C:H ratios less than 0.50(F and M, Y and M, H and V, D and K, E and R, P and R; see Figure 2a). In other words, the accumulation or reduction of an amino acid in a species can influences others in order to keep the balance of C:H ratios. We set the relationships between the C:H ratios and the amount of codon (Figure 2b), O:C ratios (Fiuge S1a) and N:C ratios (Fiuge S1b) of 20 amino acids to explore whether the distribution of 64 codons in 20 amino acids concerned the balance of C:H ratios. 20 points in Figure 2b were under red line unlike Fiuge S1a and S1b. There was a negative correlation between the the amount of codon and C:H ratios(r=-0.504,

p=0.023). The relationships between C:H ratios and O:C ratios (Fiuge 2c)and between C:H ratios and N:C ratios (Fiuge 2a) were similar with between the the amount of codon and C:H ratios, so we found that O:C ratios and N:C ratios also remained stable and were 0.50 and 0.27, respectively (see Figure 1b). The distribution of 64 codons and the choice of 20 amino acids in molecular evolution would be constrained to remain stable C:H , O:C and N:C ratios.

Homo sapiens have an estimated 20,000-25,000 genes[28]and in this paper we have collected 21,560 protein sequence from NBCI. The C:H ratios of 21,560 proteins were calculated(see Method) and their distribution diagram(mean: 0.50, 95% confidence interval: 0.5040 to 0.5044) was shown in Figure 3a. A leptokurtosis appeared above normal distribution curve(red line, Skewness=0.103±0.017, Kurtosis=2.498±0.033), which illustrated the C:H ratios of proteins tend to be 0.50. Skewness and Kurtosis of 33 species were shown in Figure 3b and 3c. The absolute value of their skewness were less than 0.6(<0.6) while kurtosis of 15 species were >1 and of all species were >0.25. When the absolute value of both skewness and kurtosis are <1, the date fits normal distribution(17 species).

The expression level and functional importance of proteins were reported to be a major determinant of the rate of protein sequence evolution[5]. However, little attention was given to the relationship between 20 amino acids that are composition of thousands and thousands of proteins. Based on the above findings, 12540 pearson correlations (20×19×33) between the constituent ratios of 20 amino acids of all proteins respectively in 33 species were obtained and drawn into a Heat Map(Figure 4e). 4 common relationships in Homo sapiens were shown in Figure 5a-b, such as positive correlation(r>0.3, p<0.001,Figure 5a) and negative correlation(r<-0.3, p<0.001, Figure 5b), weak correlation(|r|<0.3, p<0.001, Figure 5d), and indifference correlation(p>0.001, Figure 5c). We found there were similar relationships between amino acids in all 33 species(yellow or blue lines in Figure 4e), especially, in closely related one. In Figure 4f, there were relationships between Asp(D) and other 19 amino acids and we found that Glu(E) was positive correlation to it in all 33 species(yellow squares in Figure 4f). The neutral theory asserts that the vast majority of intraspecific polymorphisms and interspecific differences in protein sequence are selectively neutral rather than adaptive[29,30], which conflict with our findings. In fact, the ratios of amino acids were correlated to others in both the species and protein level. It was a pity that we had not found the direct evidences support their interactions in the latter were also related to keep the balance of the C:H ratios.

For example, the number of Gly(G) in the proteins of human was 31.98± 37.11(from 0 to 1483), and its histogram was shown Figure S2a. The length of proteins in human was 477.3± 481.8 and its histogram was shown in Figure S2b. Kolmogorov-Smirnov (KS) test was performed to verify whether those data fitted normal distribution or Poisson distribution. We found the two sets of data did not match the above distribution(all Z[?]27.84, P <0.000). We set a model that the proteins were supposed to 20 difference figures which were corresponding to the number of 20 amino acids in this protein(seeing Figure 5a and Method) and we counted the frequency of the figures(0 to 35) in every proteins of 33 species and the mean frequency of 0 to 35 in 33 species were shown in Figure 5b. The most frequently is 4, 5 and 6(vertex coordinate)and their mean values is from 0.62(Serinus canaria) to 1.25(Bacillus anthracis). On the right of vertex coordinate, when the figure is more than 10, their mean values become similary or equivalent in 33 species. When the actual mean values equal to theoretical value(λmean=λ), the data fits Poisson distribution(a random event), which was shown in Figure 5c. In Figure 5d, the frequencies of figures(6 and 20) in human were shown and we found that the mean of the figure(6) repeated two, three and four times in a protein increased while the mean of the figure(20) repeated two, three and four times was similar with the theoretical value(Poisson distribution).

To find and explore the similarities of species in evolution maybe a way of solving the mystery of origin of organisms. In 20th century, the continued perfection of genetic central dogma is an example. It is very difficult to find the the similarities of species in the base and protein sequences because of biodiversity. In the past 15 years, the increased availability of genomic data for species and biological information technology offer favorable conditions. In this paper, we study interrelation between 20 amino acid through the collection of of genomic data of 33 different species. Three similarities of species as a result of constraints between amino acids in evolution were found, such as relatively stable carbon and hydrogen ratios, their interactions

and Poisson distribution. These constraints are conducive to understand 64 codons are unequally assigned to 20 amino acids, 20 amino acids selected to composition of proeteins, and the protein sequence evolution.

Methods

All 33 species of genomes were extracted from the NCBI genome database (*http://www.ncbi.nlm.nih.gov/genome/*), such as Pan troglodytes, Gorilla gorilla, Chlorocebus sabaeus, Homo sapiens, Serinus canaria, Ovis aries, Sus scrofa, Mus musculus, Gorilla gorilla, Drosophila melanogaster, Gallus gallus, Danio rerio, Zea mays,Bordetella pertussis, Bordetella bronchiseptica, Pseudomonas putida, Pseudomonas aeruginosa, Pseudomonas syringae, Escherichia coli, Salmonella enterica, Vibrio cholerae, Vibrio parahaemolyticus, Saccharomyces cerevisiae, Arabidopsis thaliana, Helicobacter pylori, Bacillus anthracis, Streptococcus pneumoniae, Staphylococcus aureus, Staphylococcus epidermidis, Bacillus subtilis, Streptococcus pyogenes, Chlamydia trachomatis, and Buchnera aphidicola. Then we removed repeating gene sequences and collected the Protein product (NCBI Reference Sequence: NP_ or XP_).

The sequence of proteins in 33 species of genomes was collected via NCBI Reference Sequence from the NCBI protein database (*http://www.ncbi.nlm.nih.gov/protein/*). The number of 20 amino acids in every proteins were calculated and analyzed by statistics. According to neutral theory, mutations as kind of a random event, the rates of 20 amino acid in a species should be similar with the rates of their corresponding codons and their difference was denote by:

$F=\frac{n*61}{N*c}-1$

Where N is the sum of 20 amino acids in the genome of a species; n is the number of one amino acid in the genome of a species; and c is the number of their corresponding codons. 61 is the sum of codons(exclude 3 termination codons).

The number of carbon was by:

$C=2\times G+3\times A+5\times V+6\times L+6\times I+9\times F+11\times W+9\times Y+4\times D+4\times N+5\times E+6\times K+5\times Q+5\times M+3\times S+4\times T+3\times C+5\times P+6\times H$

Where the abbreviation of 20 amino acid is their number in a protein, such as G:$C_2H_5NO_2$, A:$C_3H_7NO_2$, V:$C_5H_{11}NO_2$, L:$C_6H_{13}NO_2$, I:$C_6H_{13}NO_2$, F:$C_9H_{11}NO_2$, W:$C_{11}H_{12}N_2O_2$, Y:$C_9H_{11}NO_3$, D:$C_4H_7NO_4$, N:$C_4H_8N_2O_3$, E: $C_5H_9NO_4$, K:$C_6H_{14}N_2O_2$, Q:$C_5H_{10}N_2O_3$, M:$C_5H_{11}O_2NS$, S:$C_3H_7NO_3$, T:$C_4H_9NO_3$, C:$C_3H_7NO_2S$, P:$C_5H_9NO_2$, H($C_6H_9N_3O_2$, R:$C_6H_{14}N_4O_2$. The number of hydrogen, oxygen, nitrogen were counted using the same method.

In this paper, we set a model that every protein was supposed to be made from 20 figures which is the number of 20 amino acids(see Figure 5a). The figures are from 0 to 33. The frequency of those figures appearing in a protein were calculated in 33 species to find their similarities under what conditions they accords with Poisson distribution.

Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time and/or space if these events occur with a known average rate and independently of the time since the last event. The probability of observing k events in an interval is given by the equation:

$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$

$P(X=k)=\frac{\lambda}{k}P(X=k-1)$

$P(X=0)+P(X=1)+P(X=2)+P(X=3)+..+P(X=k)=1$

Where λ is the average number of events per interval; e is the number 2.71828... (Euler's number) the base of the natural logs; k takes values 0, 1, 2, ...; and k! is the factorial of k = k × (k - 1) x (k - 2) x ... x 2 x 1.

When $\lambda_{mean}=\lambda=\frac{P(X=k)}{P(X=k-1)}k$, it accords with Poisson distribution.

λ_{mean}= P(X=0) ×0+ P(X=1) ×1+ P(X=2) ×2+ P(X=3) ×3+..+ P(X=k) ×k

In this paper, all statistical analysis was performed with SPSS 19.0. The bivariate correlation was analyzed via pearson correlation coefficient and normal and Poisson distribution test were used Kolmogorov-Smirnov (KS) test.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 81603016, 81773624), the Natural Science Foundation of Jiangsu Province (No. BK20160706, BE2017746), the National Science and Technology Major Project (2018ZX09301026-005, 2020ZX09201015).

Additional Information

The authors declare no competing financial interests.

Data Accessibility Statement

All data were extracted from the NCBI database (*http://www.ncbi.nlm.nih.gov/*)

References

1 Kimura, M. Evolutionary rate at the molecular level. *Nature* **217** , 624-626 (1968).

2 Kumar, S. Molecular clocks: four decades of evolution. *Nature reviews. Genetics* **6** , 654-662, doi:10.1038/nrg1659 (2005).

3 Qian, W., He, X., Chan, E., Xu, H. & Zhang, J. Measuring the evolutionary rate of protein-protein interaction. *Proceedings of the National Academy of Sciences of the United States of America* **108** , 8725-8730, doi:10.1073/pnas.1104695108 (2011).

4 Takahata, N. Molecular clock: an anti-neo-Darwinian legacy. *Genetics* **176** , 1-6, doi:10.1534/genetics.104.75135 (2007).

5 Zhang, J. & Yang, J. R. Determinants of the rate of protein sequence evolution. *Nature reviews. Genetics* **16** , 409-420, doi:10.1038/nrg3950 (2015).

6 Thompson, J. F., Morris, C. J. & Smith, I. K. New naturally occurring amino acids. *Annual review of biochemistry* **38** , 137-158, doi:10.1146/annurev.bi.38.070169.001033 (1969).

7 Martincorena, I., Seshasayee, A. S. & Luscombe, N. M. Evidence of non-random mutation rates suggests an evolutionary risk management strategy. *Nature* **485** , 95-98, doi:10.1038/nature10995 (2012).

8 Cook, H. Evolution For Everyone: How Darwin's Theory Can Change the Way We Think about Our Lives.*Quarterly Review of Biology* **84** , 96-97 (2009).

9 Bromham, L. & Penny, D. The modern molecular clock. *Nature reviews. Genetics* **4** , 216-224, doi:10.1038/nrg1020 (2003).

10 Worth, C. L., Gong, S. & Blundell, T. L. Structural and functional constraints in the evolution of protein families. *Nature reviews. Molecular cell biology***10** , 709-720, doi:10.1038/nrm2762 (2009).

11 Crick, F. H. The origin of the genetic code. *Journal of molecular biology* **38** , 367-379 (1968).

12 Woese, C. R., Dugre, D. H., Dugre, S. A., Kondo, M. & Saxinger, W. C. On the fundamental nature and evolution of the genetic code. *Cold Spring Harbor symposia on quantitative biology* **31** , 723-736 (1966).

13 Wong, J. T. A co-evolution theory of the genetic code. *Proceedings of the National Academy of Sciences of the United States of America* **72** , 1909-1912 (1975).

14 P., X. Critical Reviews and Reconstruction of Evolutionary Theories. . *Beijing: Science Press (in Chinese)* (2016).

15 Dos Reis, M., Donoghue, P. C. & Yang, Z. Bayesian molecular clock dating of species divergences in the genomics era. *Nature reviews. Genetics* **17** , 71-80, doi:10.1038/nrg.2015.8 (2016).

16 Margoliash, E. Primary Structure and Evolution of Cytochrome C. *Proceedings of the National Academy of Sciences of the United States of America* **50** , 672-679 (1963).

17 Doolittle, R. F. & Blombaeck, B. Amino-Acid Sequence Investigations of Fibrinopeptides from Various Mammals: Evolutionary Implications. *Nature* **202** , 147-152 (1964).

18 Doolittle, R. F., Feng, D. F., Tsang, S., Cho, G. & Little, E. Determining divergence times of the major kingdoms of living organisms with a protein clock. *Science***271** , 470-477 (1996).

19 Langley, C. H. & Fitch, W. M. An examination of the constancy of the rate of molecular evolution. *Journal of molecular evolution* **3** , 161-177 (1974).

20 Felsenstein, J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of molecular evolution* **17** , 368-376 (1981).

21 Drummond, A. J., Ho, S. Y., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS biology* **4** , e88, doi:10.1371/journal.pbio.0040088 (2006).

22 Rannala, B. & Yang, Z. Inferring speciation times under an episodic molecular clock. *Systematic biology* **56** , 453-466, doi:10.1080/10635150701420643 (2007).

23 Cnossen, I. *et al.* Habitat of early life: Solar X-ray and UV radiation at Earth's surface 4–3.5 billion years ago. *Journal of Geophysical Research Atmospheres***112** , 21–24 (2007).

24 Fay, J. C. & Wu, C. I. Sequence divergence, functional constraint, and selection in protein evolution. *Annual review of genomics and human genetics* **4** , 213-235, doi:10.1146/annurev.genom.4.020303.162528 (2003).

25 Cheng, G., Qian, B., Samudrala, R. & Baker, D. Improvement in protein functional site prediction by distinguishing structural and functional constraints on protein family evolution using computational design. *Nucleic acids research***33** , 5861-5867, doi:10.1093/nar/gki894 (2005).

26 Savill, N. J., Hoyle, D. C. & Higgs, P. G. RNA sequence evolution with secondary structure constraints: comparison of substitution rate models using maximum-likelihood methods. *Genetics* **157** , 399-411 (2001).

27 Jordan, I. K. *et al.* A universal trend of amino acid gain and loss in protein evolution. *Nature* **433** , 633-638, doi:10.1038/nature03306 (2005).

28 International Human Genome Sequencing, C. Finishing the euchromatic sequence of the human genome. *Nature* **431** , 931-945, doi:10.1038/nature03001 (2004).

29 Kimura, M. The neutral theory of molecular evolution and the world view of the neutralists. *Genome / National Research Council Canada = Genome / Conseil national de recherches Canada* **31** , 24-31 (1989).

30 Kimura, M. The neutral theory of molecular evolution: a review of recent evidence. *Idengaku zasshi***66** , 367-386 (1991).

Figure 1 **The mean of carbon and hydrogen ratios keep a relatively stable value(0.50) in 33 species regardless of the significant change rates of 20 amino acids. a** , Relationships between the sum of 20 amino acids of all proteins in 33 species and the number of their corresponding codons(F=$\frac{n*61}{N*c}-1$), N: the sum of 20 amino acid in the genome of a species; n: the number of one amino acid in the genome of a species; c:the number of their corresponding codons. 61 is the sum codons. **b** , the mean of carbon and hydrogen ratios(0.50), oxygen and carbon ratios(0.50), nitrogen and carbon ratios(0.27) in 33 species.

Figure 2 **The reasons of the relatively stable mean of carbon and hydrogen ratios.** Relationships between 20 amino acids in 33 species(a), between the amout of codons and C:H ratio in 20 amino acids (b), between O:C and C:H ratio in 20 amino acids(c) and between N:C and C:H ratio in 20 amino acids (d).

Figure 3 **The carbon and hydrogen ratios of all protiens in 33 species were tend to 0.50. a** , The histogram of C:H ratio of 21560 proteins in homo sapiens(red line: normal curve); **b** , the kurtosis(red line: y=1) of the histogram of C:H ratio of 33 species d;**c** , the skewness of the histogram of C:H ratio of 33 species.

Figure 4 **The interrelations between 20 amino acids in 33 speices. a-d** , four kinds of scatter diagrams between in homo sapiens(positive correlation: r>0.3,p<0.001; negative correlation: r<-0.3,p<0.001; weak correlation: |r|<0.3,p<0.001 and indifference correlation: p>0.001);**e** , Pearson correlation analysis between 20 amino acids in 33 spacies and 12540 correlation coefficient(r, 19x20x33) were obtained and shown in the matrix, and a phylogenetic tree was below; **f** , one of the matrix(**e**): Pearson correlation analysis between Asp(D) and other 19 amino acids and the matrix of 627 correlation coefficient.

Figure 5 **Poisson distribution. a** , the model to count the frequency of the figures which is the number of 20 amino acids in a protein; **b** , the mean frequency of the figure(λ_{mean}) in 33 species, and the highest λ_{mean} was 4, 5 or 6 in 33 species; **c** , the relationship between λ_{mean} and the theoretical frequency of the figures(λ) when these figures would fit Poisson distribution, red line: y=x; **d** , the histogram of the frequency of 6 and 20 and their fitted curve under the conditions of λ_{mean} and λ.

Figure 1

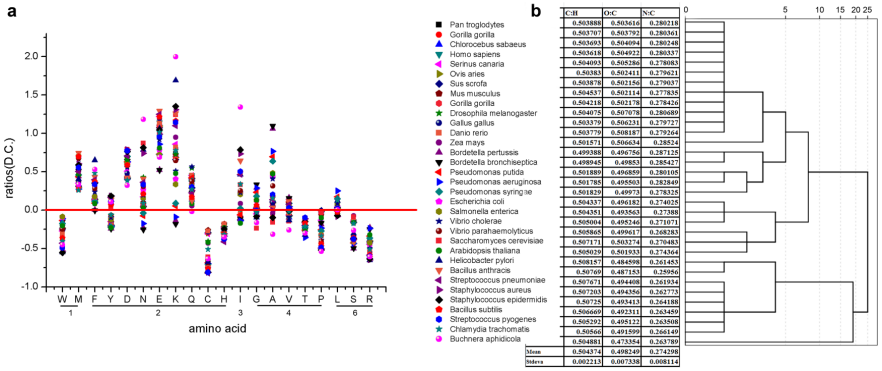

Figure 2

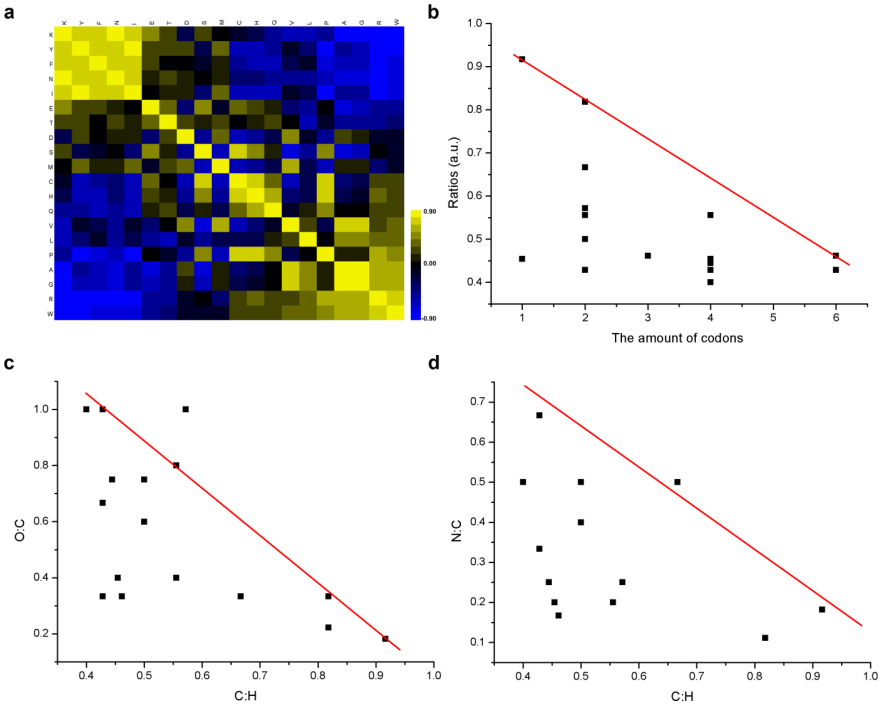

Figure 3

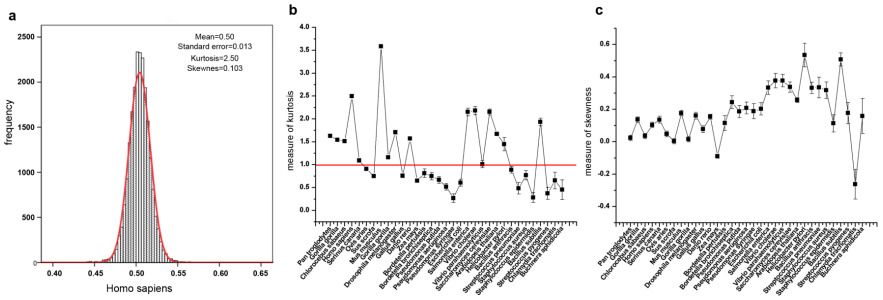

Figure 4

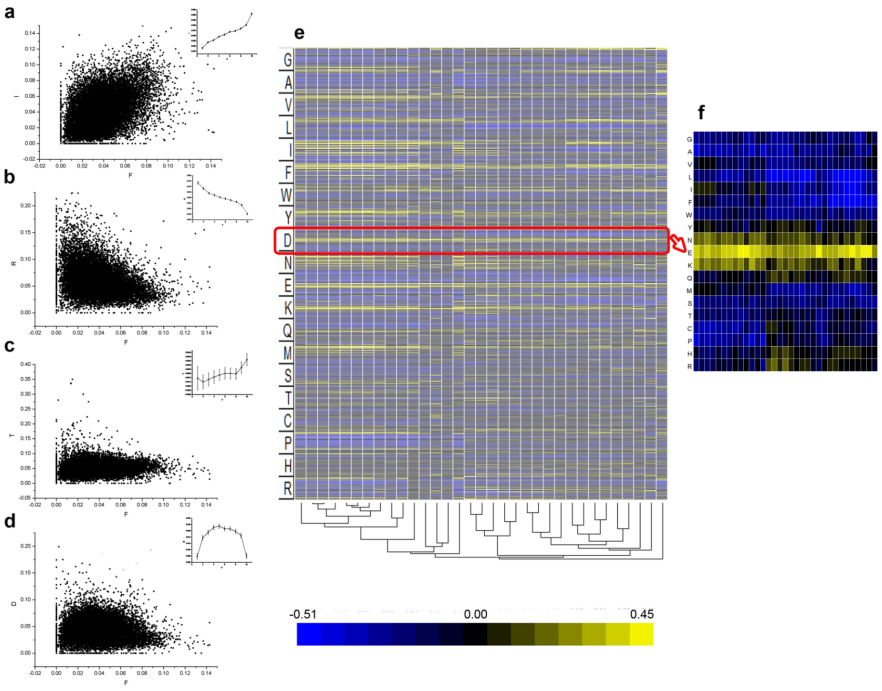

Figure 5