

RESEARCH ARTICLE

Base composition is the primary factor responsible for the variation of amino acid usage in zebra finch (*Taeniopygia guttata*)

Yousheng Rao^{1,2*}, Zhangfeng Wang^{1,2}, Wen Luo², Wentao Sheng^{1,2}, Rendian Zhang^{1,2}, Xuewen Chai^{1,2}

1 Department of Biological Technology, Nanchang Normal University, Nanchang, Jiangxi, China, **2** Jiang Xi Province Key Lab of Genetic Improvement of Indigenous Chicken Breeds, Nanchang, Jiangxi, China)

* rys8323571@aliyun.com



OPEN ACCESS

Citation: Rao Y, Wang Z, Luo W, Sheng W, Zhang R, Chai X (2018) Base composition is the primary factor responsible for the variation of amino acid usage in zebra finch (*Taeniopygia guttata*). PLoS ONE 13(12): e0204796. <https://doi.org/10.1371/journal.pone.0204796>

Editor: Suzannah Rutherford, Fred Hutchinson Cancer Research Center, UNITED STATES

Received: March 12, 2018

Accepted: September 15, 2018

Published: December 5, 2018

Copyright: © 2018 Rao et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: Funded by National Nature Science Foundation of China (31460595), Major Projects of Jiangxi Provincial Science and Technology (2015ACF60019), Jiangxi Provincial Science and Technology Landing Project (KJLD14102), and Science Technology Program of Jiangxi Education Department (GJJ14781).

Abstract

In the present study, we carried out an examination of the amino acid usage in the zebra finch (*Taeniopygia guttata*) proteome. We found that tRNA abundance, base composition, hydrophobicity and aromaticity, protein second structure, cysteine residue (Cys) content and protein molecular weight had significant impact on the amino acid usage of the zebra finch. The above factors explained the total variability of 22.85%, 25.37%, 10.91%, 5.06%, 4.21%, and 3.14%, respectively. Altogether, approximately 70% of the total variability in zebra finch could be explained by such factors. Comparison of the amino acid usage between zebra finch, chicken (*Gallus gallus*) and human (*Homo sapiens*) suggested that the average frequency of various amino acid usage is generally consistent among them. Correspondence analysis indicated that base composition was the primary factor affecting the amino acid usage in zebra finch. This trend was different from chicken, but similar to human. Other factors affecting the amino acid usage in zebra finch, such as isochore structure, protein second structure, Cys frequency and protein molecular weight also showed the similar trends with human. We do not know whether the similar amino acid usage trend between human and zebra finch is related to the distinctive neural and behavioral traits, but it is worth studying in depth.

Introduction

Amino acids are utilized with different frequencies in various proteins and organisms. Such biases in amino acid usage have been demonstrated extensively in prokaryote and eukaryote genomes, and likely reflect a balance or near balance between the action of mutation, selection, and genetic drift [1–3]. Base composition in a number of species has been shown to correlate with the amino acid content of proteins. This trend has been attributed to the neutral processes or mutation [1, 4–11]. Using a measure based on tRNA-gene copy numbers as a rough estimate of tRNA abundance, a positive correlation between tRNA abundance and the amino acid content has been documented in many organisms, suggesting selection plays an important role in shaping amino acid frequencies [3, 12–17]. In addition, intragenomic analyses have suggested

Competing interests: The authors have declared that no competing interests exist.

that factors like hydrophobicity, aromaticity, cysteine residue (Cys) content, gene function, metabolic cost, mean molecular weight and gene expression level also have significant impact on the global amino acid composition of each species [1–3, 18–23].

Although there are many influencing factors, the base composition was considered as the driving force in the amino acid usage. Knight et al. [1] made a comparative study on the impact of GC content on codon usage and amino acid usage for bacteria, archaea and eukaryotes with limited gene sample. They concluded that amino acid responses were determined by the mean GC content of their codons (explaining 71–79% of the variance). However, Rao et al. [3] reported that only approximately 10–40% variation of amino acid usage could be explained by GC content in chicken. A recent study argued that the impact could be also in the opposite direction, i.e. the selection at the amino acid level could affect the nucleotide content and codon usage significantly [24].

In the avian group, Rao et al. [3] made a systematic study of the amino acid usage in the chicken proteome. They found that the relative amino acid usage was strongly correlated with the tRNA abundance. Correspondence analysis also suggested that the main factors responsible for the variation of amino acid usage in chicken were hydrophobicity, aromaticity, and genomic GC content. In the present study, we carried out an examination of the amino acid usage in the zebra finch (*Taeniopygia guttata*) proteome. The aim of this study is to explore which are the main parameters that shape the global amino acid usage in the zebra finch, to assess the similarities and differences between the two bird genomes, and to describe their biological implications.

Materials and methods

Sequence data

In this study, gene sequences, coding DNA sequences (CDSs), or complete mRNA sequences corresponding to all annotated genes in *Taeniopygia guttata* genome were downloaded from Ensembl. For this data collection, a strict criteria was defined: (1) Only nuclear genes with known protein products (rather than a novel or predicted transcripts) were included; (2) Only genes with complete CDSs were included; (3) Genes with a CDS that did not begin with an ATG start codon, or did not have a length ≥ 300 bp, or did not occur in multiples of three nucleotides, or contained an internal stop codon, were discarded. We declare that all the data used in this study are public.

tRNA gene copy number data

The copy numbers of individual tRNA genes in the *Taeniopygia guttata* genome were taken from (<http://gtrnadb.ucsc.edu/GtRNAdb2/genomes/eukaryota/Tgutt2/>). In this data set, pseudogenes have already been removed.

Correspondence analysis

Correspondence analysis (COA) implemented by CodonW 1.4.2 was used to identify the major factors that shape variation in amino acid usage among proteins of *Taeniopygia guttata*. For each gene, the relative amino acid usage (RAAU), the GC content of the CDS (GC_{cds}), the GC content at the first, the second and the third position ($GC1$, $GC2$ and $GC3$), the average hydrophobicity (general average hydrophobicity, GRAVY) and the average aromaticity (average aromaticity, Aromo), were calculated by CodonW 1.4.2. We also performed a principal-components analysis (PCA) for the genes of zebra finch. The results were similar to the correspondence analysis.

Statistical analysis

Correlation analysis between variables was performed by SAS Proprietary Software Release 8.1. In order to assess the actual strength of correlation, all correlation coefficients reported in this study were tested independently, excluding the influence of other related variables. To determine the variables contributing to the variability and how they may interact, we performed multiple linear regressions with the variables, excluding those not contributing significantly through the use of the t-statistical logarithm with backward stepwise regression. The significance tests were corrected for multiple testing by the Bonferroni step-down correction [25].

Results

Relationship between amino acid usage and tRNA gene copy number

The relative amino acid usage (RAAU) for each gene was calculated by CodonW 1.4.2. We found that the amino acids were not equally used in the zebra finch proteome. The average RAAU of Leu, Ser, Ala, Lys, Glu and Arg was relatively high (> 6%); otherwise, some amino acid RAAU such as Cys, His, Trp, Tyr was relatively low (< 3%). We retrieved the tRNA gene copy numbers for each codon in the *Taeniopygia guttata* genome. The isoaccepting tRNA genes were summed for each amino acid. Our data demonstrated that the average RAAU was correlated with the isoaccepting tRNA gene copy numbers significantly ($r = 0.478$, $p = 0.038$) (Fig 1).

Factorial correspondence analysis for amino acid usage

Correspondence analysis (COA) was used to explore the major factors shaping variation in amino acid usage among *Taeniopygia guttata* proteins. The coordinate of each gene on each axis and the fraction of the total variation accounted for by each axis was generated by COA. Our data indicated that 4 of the 19 axes accounted for almost 50% of the total variance (47.79%) in amino acid composition of *Taeniopygia guttata* proteins. The first major axis accounted for 17.34% of the total variance, and the 2nd, 3rd, 4th major axis accounted for

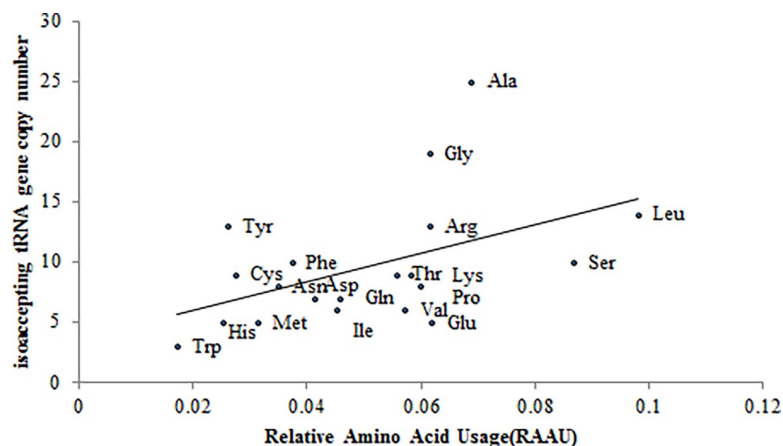


Fig 1. Relationship between the relative amino acid usage and the isoaccepting tRNA gene copy number. The tRNA gene copy numbers for each codon in the *Taeniopygia guttata* genome was taken from <http://gtrnadb.ucsc.edu/GtRNAdb2/genomes/eukaryota/Tgutt2/> (August 2, 2017). The isoaccepting tRNA gene number was summed for each amino acid. The relative amino acid usage (RAAU) for each amino acid was calculated by CodonW 1.4.2. The average RAAU values of amino acid was correlated with the isoaccepting tRNA gene copy numbers significantly ($r = 0.478$, $p = 0.038$).

<https://doi.org/10.1371/journal.pone.0204796.g001>

14.68%, 8.5%, 7.27% of the total variance, respectively. The distribution of the amino acid residues and the total genes for the first two axes were shown in Fig 2.

Impact of GC content on amino acid usage

The GC content of the CDS (GC_{cds}), the GC content at the first, the second and the third position (GC1, GC2 and GC3), were calculated by codonW 1.4.2. As shown in Fig 3, axis 1 was positively correlated with the GC_{cds} , GC2, and GC3, significantly (axis 1 vs. GC_{cds} , $r = 0.543$, $p < 0.0001$; axis 1 vs. GC2, $r = 0.887$, $p < 0.0001$; axis 1 vs. GC3, $r = 0.186$, $p < 0.0001$). Multiple regression analysis indicated that the main factor was GC2 ($R^2 = 0.788$, $p < 0.0001$). Axis 2 was negatively correlated with GC_{cds} , GC1, and GC2, significantly (axis 2 vs. GC_{cds} , $r = -0.427$, $p < 0.0001$; axis 2 vs. GC1, $r = -0.343$, $p < 0.0001$; axis 2 vs. GC2, $r = -0.517$, $p < 0.0001$). Multiple regression analysis indicated that the main factors were GC2 and GC1 ($R^2 = 0.425$, $p < 0.0001$). Axis 3 and axis 4 also correlated with the GC content. The main factors for Axis 3 were GC1 and GC2 ($R^2 = 0.556$, $p < 0.0001$), while the main factors affecting Axis 4 were GC1 and GC_{cds} ($R^2 = 0.126$, $p < 0.0001$). According to Sabbia *et al.* [20], we used the GC content of the surrounding regions of gene (25 kb upstream of the initiation codon plus the 25 kb downstream of the stop codon) as an estimator for isochores structure, and made correlation analysis between the estimator and axis 1, axis 2. We found that both axis 1 and axis 2 were significantly

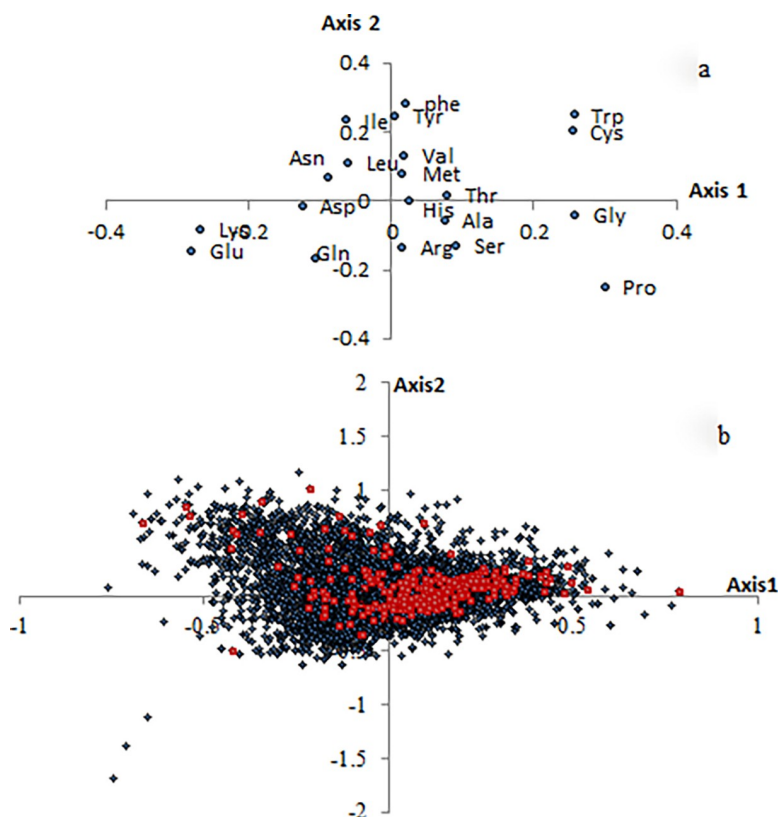


Fig 2. Distribution of the amino acids and genes on the first two axes of the correspondence analysis. a. Representation of the first two axes of the correspondence analysis performed on the amino acid frequency of *Taeniopygia guttata* protein. b. Representation of the first two axes of the correspondence analysis performed on the amino acid frequencies of 8109 *Taeniopygia guttata* genes. Membrane proteins are indicated by red dots. The total number of membrane proteins was 298. The percentage of membrane proteins with the positive value account for 72%.

<https://doi.org/10.1371/journal.pone.0204796.g002>

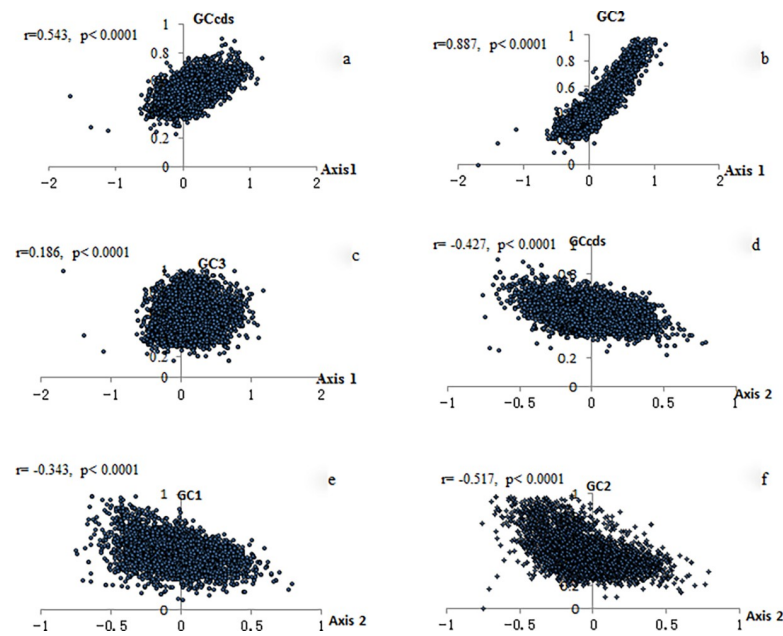


Fig 3. Relationship between GC content and axis 1,axis 2. a. Axis 1 positively correlated with GC_{cds} significantly. b. Axis 1 strongly correlated with GC2 positively. c. Axis 1 weakly correlated with GC3 positively. d. Axis 2 negatively correlated with GC_{cds} significantly. e. Axis 2 negatively correlated with GC1 significantly. f. Axis 2 negatively correlated with GC2 significantly.

<https://doi.org/10.1371/journal.pone.0204796.g003>

correlated with the estimator, suggesting that the isochore structure had a significant impact on the amino acid usage in zebra finch (axis 1 vs. estimator, $r = 0.198, p < 0.0001$; axis 2 vs. estimator, $r = -0.16, p < 0.0001$).

Impact of hydrophobicity and aromaticity on amino acid usage

The average hydrophobicity (general average hydrophobicity, GRAVY) and aromaticity (average aromaticity, Aromo), were calculated by CodonW 1.4.2. As shown in Fig 4, axis 2 was strongly correlated with the GRAVY score of proteins ($r = 0.732, p < 0.0001$), and the Aromo score of proteins ($r = 0.689, p < 0.0001$). As axis 2 was also found to be correlated with GC_{cds}, GC1, and GC2 significantly, we made a multiple regression analysis between axis 2 and all 5 variables. Our data indicated that 90.4% of the total variation of axis 2 could be explained by GRAVY, Aromo, GC2 and GC1 ($R^2 = 0.904, p < 0.0001$), and 74.3% could be explained by Gravy and Aromo ($R^2 = 0.743, p < 0.0001$). Axis 1 also showed a significant correlation with the GRAVY score ($r = 0.228, p < 0.0001$), but did not correlate with the Aromo score ($r = 0.01, p = 0.334$). As shown in Fig 2A, the strong hydrophobic amino acid Ile, Val, Phe, Leu, Met, and the aromatic amino acid Tyr, Phe, and Trp, were at the above of the plane (positive values for axis 2). The distribution of genes in Fig 2B indicated that the membrane proteins were related to the distribution of axis 2, in which the majority of them showed a positive value over the axis 2.

Impact of protein second structure, molecular weight and Cys frequency on amino acid usage

The amount of secondary structure for each protein was predicted by the use of PHD software. The distribution of alpha helix, extend strand, and random coil over the entire protein data set were analyzed. Correlation analyses indicated that extend strand was positively correlated with

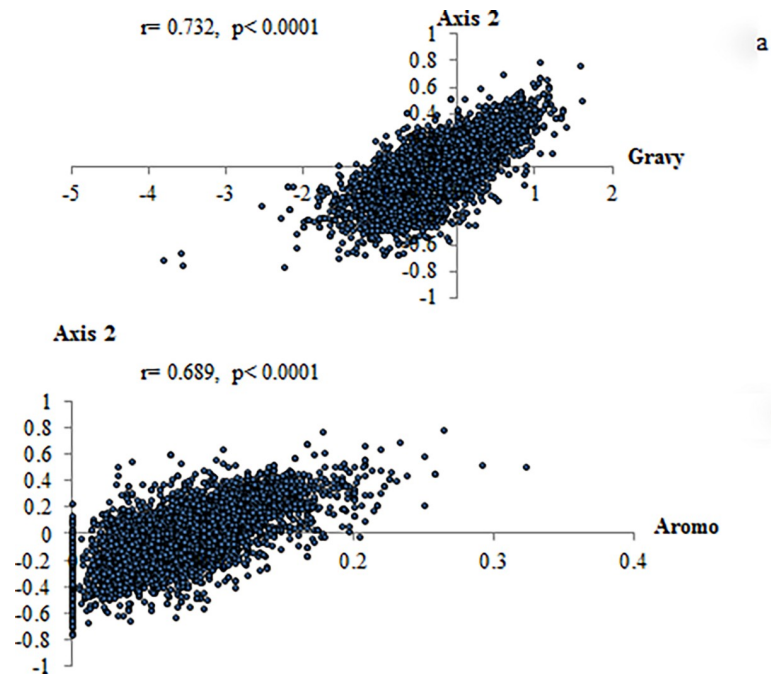


Fig 4. Relationship between axis 2 and the GRAVY score of proteins, the Aromo score of proteins. a. Axis 2 strongly correlated with the GRAVY score of proteins. b. Axis 2 strongly correlated with the Aromo score of proteins.

<https://doi.org/10.1371/journal.pone.0204796.g004>

axis1 ($r = 0.157$, $p < 0.0001$) and axis2 ($r = 0.354$, $p < 0.0001$). Random coil was positively correlated with axis1 ($r = 0.204$, $p < 0.0001$), negatively correlated with axis2 ($r = -0.229$, $p < 0.0001$) and axis4 ($r = -0.335$, $p < 0.0001$). The second structure could explain 5.06% of the total variability found in our proteins.

Previous studies demonstrated that the molecular weight of proteins had a significant effect on the amino acid usage. The same trend was also found in *Taeniopygia guttata* proteome. The molecular weight of proteins were negatively correlated with axis 1 ($r = -0.4261$, $p < 0.0001$). We also analyzed the influence of the Cys frequency on the total variability. We found that Cys frequency was positively correlated with axis 1 ($r = 0.35$, $p < 0.0001$) and axis 3 ($r = 0.494$, $p < 0.0001$). 4.21% of the total variability could be explained by Cys frequency. The CDS length also showed a significant correlation with axis 2 ($r = -0.2179$, $p < 0.0001$) and axis 3 ($r = -0.1252$, $p < 0.0001$), but with very low coefficients.

Discussion

In the present study, we carried out a genome scale analysis of the amino acid usage in the zebra finch (*Taeniopygia guttata*). The effects of tRNA abundance, base composition, hydrophobicity and aromaticity, protein second structure, Cys frequency and protein molecular weight on amino acid usage were investigated in detail. We found that the above factors influenced the variability of amino acidic composition of the zebra finch proteome, explaining 22.85%, 25.37%, 10.91%, 5.06%, 4.21%, and 3.14% of the total variability, respectively. Altogether, approximately 70% (71.54%) of the total variability in the *Taeniopygia guttata* proteome could be explained by such factors.

Among the avian species, chicken (*Gallus gallus*) is the best studied representative. The chicken and zebra finch lineages diverged about 100 million years ago. Their genome structures are similar, such as smaller, tighter, marked reduction of interspersed repeats etc., but

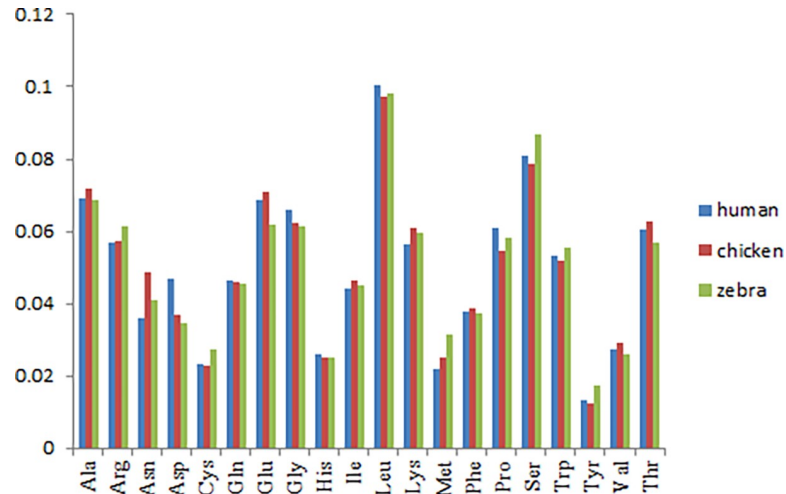


Fig 5. Comparison of the average frequency of various amino acids usage among zebra finch (*Taeniopygia guttata*), chicken (*Gallus gallus*) and human (*Homo sapiens*).

<https://doi.org/10.1371/journal.pone.0204796.g005>

they differ in many intrachromosomal rearrangements, lineage-specific gene family expansions, the number of long-terminal-repeat-based retrotransposons, and so on [26]. The zebra finch is an ideal model for study on the brain development, as it communicates through learned vocalizations, an ability documented only in human (*Homo sapiens*) and a few other animals, but lacking in chicken [27]. Here, we made a comparison of the amino acid usage among zebra finch, chicken and human. As shown in Fig 5, there was no significant difference in the average use of 20 amino acids. The trends of the various amino acid usage frequency were generally consistent. For example, the contents of Leu, Ser, Ala and Glu were relatively high, otherwise, some amino acid contents such as Cys, His, Met, Tyr were relatively low (S1 Table). Rao et al. [3] argued that the primary factors responsible for the variation of amino acid usage in chicken were hydrophobicity and aromaticity. In that study, axis 1 was strongly correlated with the GRAVY score and Aromo score. This correlation trend was not consistent with the present study. Correspondence analysis and multiple linear regression analysis in the zebra finch indicated that axis 1 was mainly influenced by GC2 ($R^2 = 0.788$, $p < 0.0001$), otherwise, hydrophobicity and aromaticity were main factors impacted on axis 2 ($R^2 = 0.743$, $p < 0.0001$). In other words, base composition was the primary factor responsible for the variation of amino acid usage in zebra finch. This trend is similar to a previous study in human [20]. Other factors affecting the amino acid usage in zebra finch, such as isochore structure, protein second structure, Cys frequency and protein molecular weight also showed the similar trends with human (S2 Table). We do not know whether the similar amino acid usage trend between human and zebra finch is related to the distinctive neural and behavioral traits, but it is worth studying in depth.

There was no significant difference in the average use of 20 amino acids among zebra finch (*Taeniopygia guttata*), chicken (*Gallus gallus*) and human (*Homo sapiens*). The trends of the various amino acid usage frequency were generally consistent.

Supporting information

S1 Table. Comparison of amino acid usage frequency among *Taeniopygia guttata*, *Gallus gallus* and *Homo sapiens*.

(DOCX)

S2 Table. Factors affecting amino acid usage in *Taeniopygia guttata*, *Gallus gallus* and *Homo sapiens*.

(DOCX)

Acknowledgments

We thank reviewers for their helpful comments on the manuscript. This work was supported by the funds from the National Nature Science Foundation of China (31460595), Major Projects of Jiangxi Provincial Science and Technology (2015ACF60019), Jiangxi Provincial Science and Technology Landing Project (KJLD14102), and Science Technology Program of Jiangxi Education Department (GJJ14781)

Author Contributions

Conceptualization: Zhangfeng Wang.

Funding acquisition: Yousheng Rao.

Investigation: Yousheng Rao, Wen Luo, Rendian Zhang, Xuewen Chai.

Project administration: Yousheng Rao.

Software: Wentao Sheng.

References

1. Knight RD, Freeland SJ and Landweber LF (2001) A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. *Genome Biol.*, 2 (4): RESEARCH0010.
2. Rispé C, Delmotte F, van Ham RC and Moya A (2004) Mutational and selective pressures on codon and amino acid usage in *Buchnera*, endosymbiotic bacteria of aphids. *Genome Res.* 14(1): 44–53. <https://doi.org/10.1101/gr.1358104> PMID: 14672975
3. Rao YS, Wang ZF, Chai XW, Nie QH and Zhang XQ (2014) Hydrophobicity and aromaticity are primary factors shaping variation in amino acid usage of chicken proteome. *PLoS One* 9(10):e110381. <https://doi.org/10.1371/journal.pone.0110381> PMID: 25329059
4. Palacios C and Wernegreen J J (2002) A strong effect of AT mutational bias on amino acid usage in *Buchnera* is mitigated at high-expression genes. *Mol. Biol. Evol.* 19(9): 1575–84. <https://doi.org/10.1093/oxfordjournals.molbev.a004219> PMID: 12200484
5. Lightfield J, Fram NR and Ely B (2011) Across bacterial phyla, distantly-related genomes with similar genomic GC content have similar patterns of amino acid usage. *PLoS One* 6(3): e17677. <https://doi.org/10.1371/journal.pone.0017677> PMID: 21423704
6. Li J, Zhou J, Wu Y, Yang S and Tian D (2015) GC-Content of Synonymous Codons Profoundly Influences Amino Acid Usage. *G3 (Bethesda)* 5(10):2027–36. <https://doi.org/10.1534/g3.115.019877> PMID: 26248983
7. Mondal SK, Kundu S, Das R and Roy S (2016) Analysis of phylogeny and codon usage bias and relationship of GC content, amino acid composition with expression of the structural *nif* genes. *J. Biomol. Struct. Dyn.* 34(8):1649–66. <https://doi.org/10.1080/07391102.2015.1087334> PMID: 26309237
8. Seligmann H (2003) Cost minimization of amino acid usage. *J. Mol. Evol.* 56: 151–161. <https://doi.org/10.1007/s00239-002-2388-z> PMID: 12574861
9. Mackiewicz P, Gierlik A, Kowalczyk M, Dudek MR and Cebrat S (1999) How does replication-associated mutational pressure influence amino acid composition of proteins? *Genome Res* 9: 409–416. PMID: 10330120
10. Mackiewicz P, Gierlik A, Kowalczyk M, Szczepanik D, Dudek MR and Cebrat S (1999) Mechanisms generating long-range correlation in nucleotide composition of the *Borrelia burgdorferi* genome. *Physica A* 273: 103–115.
11. Lafay B, Lloyd AT, McLean MJ, Devine KM, Sharp PM, Wolfe KH (1999) Proteome composition and codon usage in spirochaetes: species-specific and DNA strand-specific mutational biases. *Nucleic Acids Res.* 27(7):1642–9 PMID: 10075995

12. Du MZ, Wei W, Qin L, Liu S, Zhang AY, Zhang Y, et al. (2017) Co-adaptation of tRNA gene copy number and amino acid usage influences translation rates in three life domains. *DNA Research* 24 (6):623–633 <https://doi.org/10.1093/dnares/dsx030> PMID: 28992099
13. Kanaya S, Yamada Y, Kudo Y and Ikemura T (1999) Studies of codon usage and tRNA genes of unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. *Gene* 238:143–155. PMID: 10570992
14. Duret L (2000) tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. *Trends Genet.* 16: 287–289. PMID: 10858656
15. Behura SK, and Severson DW (2011) Coadaptation of isoacceptor tRNA genes and codon usage bias for translation efficiency in *Aedes aegypti* and *Anopheles gambiae*. *Insect. Mol. Biol.* 20(2): 177–87. <https://doi.org/10.1111/j.1365-2583.2010.01055.x> PMID: 21040044
16. Novoa EM, and Ribas de Pouplana L (2012) Speeding with control: codon usage, tRNAs, and ribosomes. *Trends Genet.* 28(11): 574–81. <https://doi.org/10.1016/j.tig.2012.07.006> PMID: 22921354
17. Qian W, Yang JR, Pearson NM, Maclean C, and Zhang J (2012) Balanced codon usage optimizes eukaryotic translational efficiency. *PLoS Genet.* 8(3): e1002603. <https://doi.org/10.1371/journal.pgen.1002603> PMID: 22479199
18. Lobry JR, and Gautier C (1994) Hydrophobicity, expressivity and aromaticity are the major trends of amino-acid usage in 999 *Escherichia coli* chromosome-encoded genes. *Nucleic Acids Res.* 22(15): 3174–80. PMID: 8065933
19. Akashi H, and Gojobori T (2002) Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc. Natl. Acad. Sci.* 99: 3695–3700. <https://doi.org/10.1073/pnas.062526999> PMID: 11904428
20. Sabbia V, Piovani R., Naya H, Rodríguez-Maseda H, Romero H and Musto H, (2007) Trends of amino acid usage in the proteins from the human genome. *J. Biomol. Struct. Dyn.* 25(1): 55–9 <https://doi.org/10.1080/07391102.2007.10507155> PMID: 17676938
21. Williford A and Demuth JP (2012) Gene Expression Levels Are Correlated with Synonymous Codon Usage, Amino Acid Composition, and Gene Architecture in the Red Flour Beetle, *Tribolium castaneum*. *Molecular Biology and Evolution* 29: 3755–3766 <https://doi.org/10.1093/molbev/mss184> PMID: 22826459
22. Whittle CA and Extavour CG (2016). Expression-Linked Patterns of Codon Usage, Amino Acid Frequency, and Protein Length in the Basally Branching Arthropod *Parasteatoda tepidariorum*. *Genome Biology and Evolution* 8:2722–36. <https://doi.org/10.1093/gbe/evw068> PMID: 27017527
23. Dufton MJ (1997) Genetic Code Synonym Quotas and Amino Acid Complexity: Cutting the Cost of Proteins? *Journal of Theoretical Biology* 187: 165–173. <https://doi.org/10.1006/jtbi.1997.0443> PMID: 9237887
24. Błażej P, Mackiewicz D, Wnętrzak M and Mackiewicz P (2017) The impact of selection at the amino acid level on the usage of synonymous codons. *G3 (Bethesda)*. 7(3):967–981 <https://doi.org/10.1534/g3.116.038125> PMID: 28122952
25. Holm S (1997) A simple sequentially rejective Bonferroni test procedure. *Scand. J. Stat.* 6: 65–70.
26. Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Künstner A et al. (2010) The genome of a songbird. *Nature* 464(7289):757–62 <https://doi.org/10.1038/nature08819> PMID: 20360741
27. Mello CV (2014) The zebra finch, *Taeniopygia guttata*: an avian model for investigating the neurobiological basis of vocal learning. *Cold Spring Harb Protoc.* 12:1237–42