COMPARISON OF CODON USAGE IN MITOCHONDRIAL GENOMES OF RHINOLOPHID AND HIPPOSIDERID BATS

Semir Dorić[1] and Lada Lukić Bilela[1,2#]

[1]Dept. of Biology, Faculty of Science, University of Sarajevo, Zmaja od Bosne 33-35, 71 000 Sarajevo, Bosnia and Herzegovina; semir.doric@gmail.com; llbilela@gmail.com
[2]Biospeleological Society in Bosnia and Herzegovina, Avde Jabučice 30, 71 000 Sarajevo, Bosnia and Herzegovina; biospeld2008@gmail.com
Corresponding author

Abstract

According to current phylogenetic hypotheses, the bats of the families Rhinolophidae and Hipposideridae are sister groups nested within the clade of Pteropodiformes. The Hipposideridae are family of bats commonly known as the Old World leaf-nose bats. While this family has long been considered as a rhinolophid subfamily Hipposiderinae, it is now more generally classified as its own family. The Hipposideridae contain 10 living genera and more than 70 species, mostly in the widespread genus *Hipposideros*. This study is an attempt to confirm a distinction between these two families by a codon usage comparison of a complete set of mitochondrial protein-coding genes from currently available mitochondrial (mt) genomes of rhinolophid and hiposiderid bats. The INCA 2.1 and GCUA 2.0 were used for the codon usage computing. Measure Independent of Length and Composition (MILC), was used to estimate the codon usage of 13 mt protein-coding genes from five species of genus *Rhinolophus* and one species of *Hipposideros* (while only four genes were available from *H. larvatus*). Large randomly generated sequence sets were used to test for dependence on (i) sequence length, (ii) overall amount of codon bias and (iii) codon bias discrepancy in the sequences. Our findings suggest no significant differences in codon usage bias, among analyzed rhinolophid species, by statistical estimation of absolute frequency values despite the changed MILC values for *nd1* and *nd3* from *Hipposideros armiger*.

Keywords: MILC, MELP, bats, codon usage, codon frequencies

1. Introduction

Rhinolophidae split from their sister family, the Hipposideridae, towards the end of the Eocene (Maree and Grant, 1997; McKenna and Bell, 1997; Teeling et al., 2003; Eick et al., 2005). This estimate of divergence is congruent with the fossil data, with fossils of extinct Rhinolophus and Hipposideros species first occurring in middle Eocene deposits (ca. 49–37 MYA; Simmons and Geisler, 1998). The family Rhinolophidae Gray, 1825 consists of a single genus *Rhinolophus* Lacépéde, 1799. The taxon is exclusively Old World, with at least 77 species (Simmons, 2005) occurring in both temperate and tropical areas throughout the Afrotropical, Australian, Indomalayan, Oceanian and Palaearctic regions (Csorba et al., 2003). A previous classifications imply two subfamilies, the Hipposiderinae and the Rhinolophinae, of the family Rhinolophidae according to Koopman, 1993, 1994; McKenna and Bell, 1997; Simmons and Geisler, 1998; Teeling et al., 2002, while contemporaneously the hipposiderids excluded from this family following Corbet and Hill (1992), Bates and Harrison (1997) and Simmons (2005). Interestingly, *Rhynolophus monoceros* (used in our study) is often treated as a Taiwanese endemic, and very similar to *R. pusillus* of the mainland in terms of body size, echolocation call frequency and mitochondrial gene sequences (Li et al. 2006). It is perhaps best treated as a synonym of *R. pusillus*, especially given that the Taiwanese *Hipposideros terasensis* considered synonymous with *H. armiger* of the mainland according to Simmons (2005).

Besides, in more recent studies of deep rhinolophid phylogeny based on analysis of cytochrome b and the three nuclear introns: thyrotropin, thyroglobulin and protein-kinase (PRKC1) the hipposiderid bats were used as outgroup (Stoffberg et al., 2010) A multidisciplinary approach: morphometric measurements (Bogdanowicz, 1992), recording and analysis of echolocation signals, karyotypic variation (Koubínová et al, 2010), D-loop sequence analysis) contribute in resolving t1he correct phylogenetic position of the species within these two families (Stoffberg et al, 2010). According to current phylogenetic hypotheses, the bats of the families Rhinolophidae and Hipposideridae are sister groups nested within the clade of Pteropodiformes (Koubínová et al, 2010). Rhinolophidae form a monophyletic group and can be divided into at least two major clades – the predominantly African and the predominantly Oriental clades – based on the current biogeographical distributions of the majority of species within each clade. Morphological (Bogdanowicz, 1992) and cytochrome b (Guillén-Servent et al., 2003) analyses also suggest that the African rhinolophids form a monophyletic clade. The typical metazoan mitochondrial (mt) genome comprises a single circular, double-stranded DNA molecule with a size between 14 and 18 kb that contains a uniform set of 37 genes (Boore, 1999). Mitochondrial genomes are powerful tool in phylogenetic analyses to elucidate the complex relationships among taxa. More rapidly evolving mitochondrial genes may distinguish even closely related species and thus they have been employed in conservation genetic studies (Avise, 1995). Thus, generation of full mt-genome sequences is important for both evolutionary studies and conservation management of endangered species.

The aim of this study was a codon usage comparison of mitochondrial genes and identification of possible differences in codon frequency values from rhinolophid and hipposiderid bats as a contribution to phylogeny elucidation between these two families.

2. Materials and methods

The nucleotide sequences of 13 (*nd1, nd2, nd3, nd4, nd4l, nd5, nd6, cox1, cox2, cytb, cox3, atp6, atp8*) mitochondrial protein-coding genes from five species of the family Rhinolophidae and *Hipposideros armiger* (Hodgson, 1835) as well as four genes (*nd1, nd2, cox1, cytb*) from *H. larvatus* (Horsfield, 1823) were obtained from GenBank (NCBI) (Table 1).

Table 1. Selected species of the families Rhinolophidae and Hipposideridae used for sequence analysis, with a mitochondrial genome accession number in GenBank database.

Selected species	Accession number
Rhinolophus ferrumequinum (Schreber, 1774)	NC_020326.1
R. pumilus K. Andersen, 1905	NC_005434.1
R. monoceros K. Andersen, 1905	NC_005433.1
R. formosae (Sanborn, 1939)	NC_011304.1
R. luctus (Temminck, 1834)	NC_018539.1
Hipposideros armiger (Hodgson, 1835)	NC_018540.1
H. larvatus (Horsfield, 1823)	JX861075.1
H. larvatus (*nd1*)	DQ888653.1
H. larvatus (*nd2*)	JQ915493.1
H. larvatus (*cox1*)	JQ365642.1
H. larvatus (*cytb*)	EU434949.1

INCA 2.1 (Supek and Vlahoviček, 2006) and GCUA 2.0 (Fuhrmann et. all., 2004) were used for codon usage computing. Measure Independent of Length and Composition (MILC), was used to estimate the codon usage of thirteen mtDNA genes of selected species within genera *Rhinolophus* and *Hipposideros*. Large randomly generated sequence sets were used to test for dependence on (i) sequence length, (ii) overall amount of codon bias and (iii) codon bias discrepancy in the sequences. MILC Based Expression Level Predictor (MELP) was used as a measurement to quantitatively predict the levels of selected mitochondrial gene expression.

3. Results and discussion

Calculations of MILC and MELP values for thirteen and four protein-coding genes of selected species, six and seven respectively, were carried out in INCA 2.1. (Table 2). Reference MILC values were used as absolute frequencies of codon usage for selected genes (F_o). Estimation of the percentage difference of codon usage bias among selected species followed trough χ^2 statistical observations as well as GCUA 2.0.

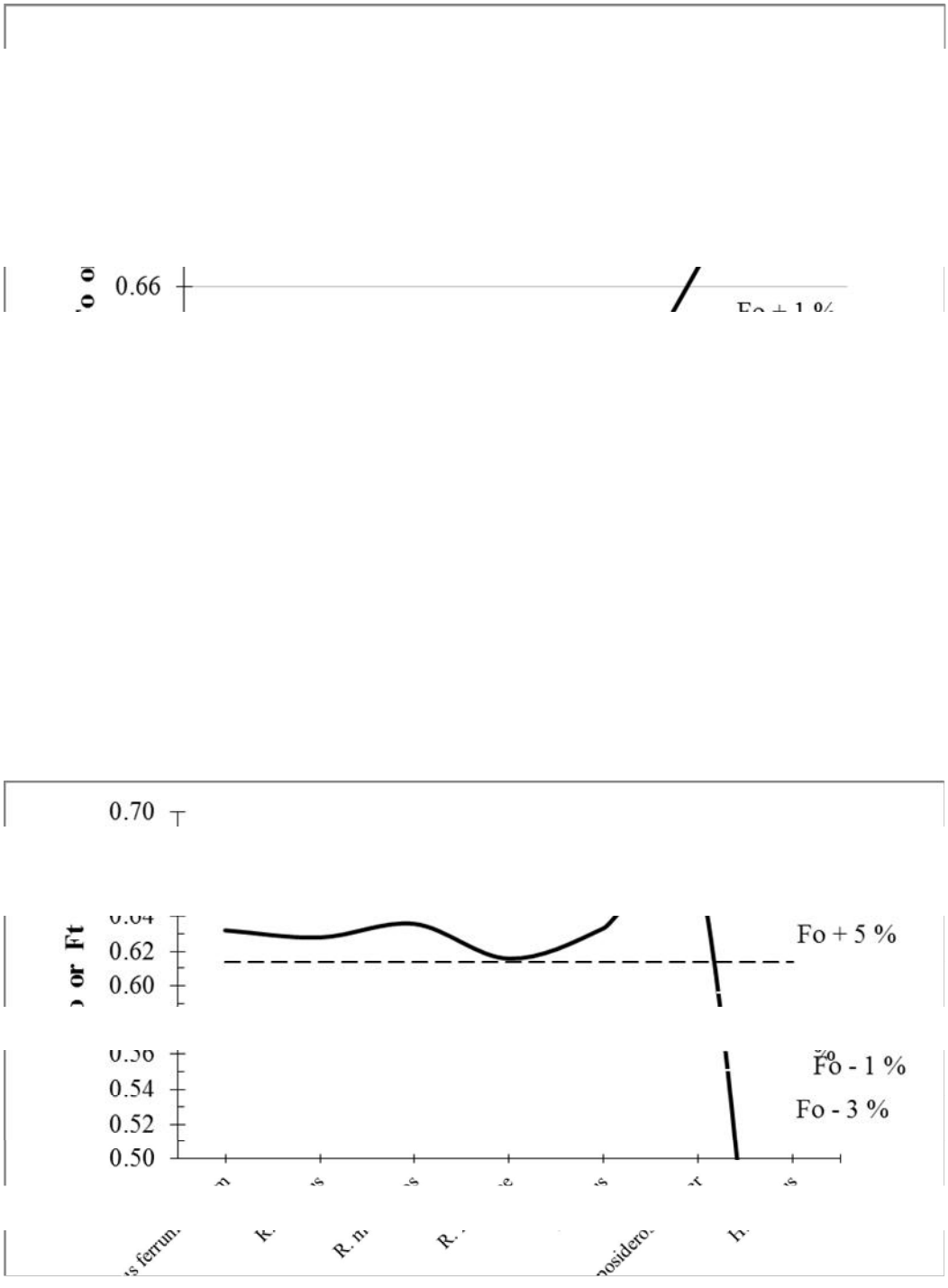
Table 2. MILC and MELP average values for 13 (from five rhinolophid and one hipposiderid bat) (A) as well as four protein coding genes from *Hipposideros larvatus* (B).

Selected species	MILC	MELP	
Rhinolophus ferrumequinum	0.631	1.011	
R. pumilus	0.627	1.009	
R. monoceros	0.635	1.015	
R. formosae	0.615	1.030	
R. luctus	0.633	1.021	
Hipposideros armiger	0.666	1.007	A

Selected species	MILC	MELP	
Rhinolophus ferrumequinum	0.516	1.057	
R. pumilus	0.525	1.050	
R. monoceros	0.527	1.054	
R. formosae	0.511	1.043	
R. luctus	0.527	1.043	
Hipposideros armiger	0.509	1.030	
Hipposideros larvatus	0.522	1.022	B

Absolute differences between the theoretical (F_t) and observed (F_o) values of codon frequencies were less than 1%, comparing thirteen and four protein-coding genes from six and seven selected species, respectively (Graph. 1).

Graph 1. Absolute difference of Ft and Fo for MILC values for 13 (from five rhinolophid and one hipposiderid bat) (A) and four protein-coding genes from *Hipposideros larvatus* (B).



The obtained value of χ^2 was 0.00218 with p=0.01 and n=5 (six species) and 0.0096 with p= 0.009 and n=6 (seven species). In both cases p value was <0.5 which indicates no significant difference. However, analysis mediated by GCUA 2.0 revealed important differences in codon usage frequencies. Absolute differences in codon usage frequency between *Hipposideros armiger* and *Rhinolophus ferrumequinum* were 8.19% , *H. armiger* and *R. monoceros* 8.17%, *H. armiger* and *R. pumilus* 9.89%, *H. armiger* and *R. luctus* 8.3%.

The typical deviation from the universal genetic code observed in mitochondrial genomes of rhinolophid and hipposiderid bats are similar to that already found in other vertebrates, with TGA coding for tryptophan, instead of being a stop codon. *Hipposideros armiger* prefers TGG (Trp) with the absolute frequency of 0.154 related to five species of genus *Rhinolophus* with frequencies of 0.05-0.65.

Among the terminations codons, UAA was the most preferred by all analyzed species, then AGA while the codon AGG is found neither in mitochondrial genes of five rhinolophid species nor in *H. armiger* and *H. larvatus*. AGA and AGG were thought to have become mitochondrial stop codons early in vertebrate evolution (Osawa, et al., 1989). However, at least in humans it has now been shown that AGA and AGG sequences are not recognized as termination codons. UAG codon is rather preferred by *H. armiger* than rhinolophid bats with a difference of 4.04%. Actually, the UGA Stop-to-Trp is the change is the most frequently occurring reassignment known. Disappearance of UAG would be favored by mutation pressure increasing the AU content. The reason UAG is reassigned less frequently than UGA may be because of the relative difficulty of the required change in the tRNA. In the case of UGA, the existing tRNA-Trp can simply mutate its anticodon (Sengupta et al., 2007).

Synonymous codons are not used with equal frequencies, so based on a multivariate analysis of codon usage data from unicellular organisms, Grantham et al. (1980) proposed the genome hypothesis implying some relationship between codon usage and taxonomic distance. A long time ago was noticed a correlation between taxonomic divergence and the similarity of the codon dialect (Ikemura, 1985; Maruyama et al. 1986).

4. Concluding remarks

During the last decade, analyses of the mitochondrial genome became a powerful tool to resolve the phylogenetic relationships among the various eukaryotic lineages and to elucidate the early events during evolution of multicellularity. MILC and MELP algorithms have proved to be excellent tools for mitogenomic phylogeny and molecular evolution studies.
The codon usage comparison of 13 mt protein-coding genes from five species of genus *Rhinolophus* and one species of *Hipposideros* has shown 8.17 to 9.89% differences of codon frequency values using GCUA 2.0. However, our results mediated by INCA 2.1., without any statistically significant difference in codon usage among selected species could be explained by the estimation based on the only one complete mitochondrial protein-coding gene set (*H. armiger*) and four genes (from *H. larvatus*) from hipposiderid bats. Furthermore, MILC algorithm has proved to be very sensitive, discriminating genes by their length, alternative stop codons and nucleotide composition. MILC values for *nd1* and *nd3* genes from hipposiderid bats differ from genes of selected rhinolophid species due to the nucleotide composition. This could indicate possible further differences on complete mtDNA sequences which can help in elucidation of phylogenetic relationships within these two chiropteran families.

5. References

Agnarsson, I., Zambrana-Torrelio, C., M., Flores-Saldana, N., P., & May-Collado, LJ. (2011). A time-calibrated species-level phylogeny of bats (Chiroptera, Mammalia). *PLOS Currents*, *10*, 13-71.

Avise, J.C. (1995). Mitochondrial DNA polymorphism and a connection between genetics and demography of relevance to conservation. *Conservation Biology*, *9*, 686–690.

Bogdanowicz, W. (1992). Phenetic relationships among bats of the family Rhinolophidae. *Acta Theriologica*, *37*, 213–240.

Boore, J.L. (1999). Animal mitochondrial genomes. *Nucleic Acids Research*, 27, 1767–1780.

Csorba, G., Ujhelyi, P., & Thomas, N. (2003). *Horseshoe Bats of the World (Chiroptera: Rhinolophidae)*. Alana Books, Shropshire. p. 160.

Eick, G.N., Jacobs, D.S., & Matthee, C.A. (2005). A nuclear DNA phylogenetic perspective on the evolution of echolocation and historical biogeography of extant bats (Chiroptera). *Molecular Biology and Evolution*, *22*, 1869–1886.

Fuhrmann, M., Hausherr, A., Ferbitz, L., Schödl, T., Heitzer, M., & Hegemann, P. (2004). Monitoring dynamic expression of nuclear genes in *Chlamydomonas reinhardtii* by using a synthetic luciferase reporter gene. *Plant Molecular Biologie*, *55*(6), 869-81.

Grantham, R., Gautier, C., Gouy, M., Mercier, R., & Pave, A. (1980). Codon catalog usage and the genome hypothesis. *Nucleic Acids Research*, *8*, 49-62.

Guillén-Servent, A., Francis, C.M., & Ricklefs, R.E. (2003). *Phylogeny and biogeography of the horseshoe bats*. In: Csorba, G., Ujhelyi, P., & Thomas, N. (Eds.), Horseshoe Bats of the World. Alana Books, p. 160.

Ikemura, T. (1985). Codon usage and tRNA content in unicellular and multicelular organisms. *Molecular Biology and Evolution*, *2*, 13-34.

Koubínová, D., Sreepada, K. S., Koubek, P., & Zima, J. (2010). Karyotypic Variation in Rhinolophid and Hipposiderid Bats (Chiroptera: Rhinolophidae, Hipposideridae). *Acta Chiropterologica*, *12*(2), 393-400.

Li, G., Jones, G., Rossiter, S.J., Chen, S.F., Parsons, S. & Zhang, S. (2006). Phylogenetics of small horseshoe bats from East Asia based on mitochondrial DNA sequence variation. *Journal of Mammalogy*, *87*, 1234-1240.

Maree, S., & Grant, W.S. (1997). Origins of horseshoe bats (Rhinolophus, Rhinolophidae) in southern Africa: evidence from allozyme variability. *Journal of Mammalian Evolution*, *4*, 195–214.

Maruyama, T., Gojobori, T., Aota, S.I., & Ikemura, T. (1986). Codon usage tabulated from the GenBank genetic sequence data. *Nucleic Acids Research*, *14*, l51-197.

McKenna, M.C., & Bell, S.K. (1997). *Classification of Mammals above the Species Level*. Columbia University Press. p. 631.

Osawa, S., Ohama, T., Jukes, T.H., & Watanabe, K. (1989). Evolution of the mitochondrial genetic code. I. Origin of AGR serine and stop codons in metazoan mitochondria. *29*(3), 202-207.

Sengupta, S., Yang, X., & Higgs, P. G. (2007). The Mechanisms of Codon Reassignments in Mitochondrial Genetic Codes. *Journal of Molecular Evolution*, *64*, 662-688.

Simmons, N.B., & Geisler, J.H. (1998). Phylogenetic relationships of Icaronycteris, Archaeonyteris, Hassianycteris and Palaeochiropteryx to extant bat lineages, with comments on the evolution of echolocation and foraging strategies in Microchiroptera. *The Bulletin of the American Museum of Natural History*, *235*, 1–182.

Simmons, N.B. (2005). *Order Chiroptera*. In: Wilson, D.E., & Reeder, D.M. (Eds.), Mammal Species of the World: A Taxonomic and Geographic Reference, third ed. Smithsonian Institution Press, pp. 312–529.

Stoffberg, S., Jacobs, D.S., Mackie, I.J., & Matthee, C.A. (2010). Molecular phylogenetics and historical biogeography of Rhinolophus bats. *Molecular Phylogenetics and Evolution*, *54*, 1–9.

Supek, F., & Vlahoviček, K. (2005). Comparison of codon usage measures and their applicability in prediction of microbial gene expressivity. BMC *Bioinformatics*, *6*, 182.

Supek, F., & Vlahoviček, K. (2006). INCA: synonymous codon usage analysis and clustering by means of self-organizing maps. *BMC Bioinformatics*, *20*(14), 2329-2330.

Teeling, E.C., Madeson, O., Van Den Bussche, R.A., De Jong, W.W., Stanhope, M.J., & Springer, M.S., (2002). Microbat paraphyly and the convergent evolution of a key innovation in Old World rhinolophoid microbats. *PNAS USA*, *99*, 1431–1436.

Teeling, E.C., Madsen, O., Murphy, W.J., Springer, M.S., & O'Brien, S.J. (2003). Nuclear gene sequences confirm ancient link between New Zealand's short-tailed bat and South American noctilionoid bats. *Molecular Phylogenetic and Evolution*, *28*, 308–319.

Vaughan, T., Ryan, J., & Czaplewski, N. (2000). *Mammalogy*, 4th Edition. Toronto: Brooks Cole Press.

Wilson, D., & Reeder, D. (2005). *Mammal Species of the World*, 3rd edition. Baltimore: Johns Hopkins University Press.

Semir Dorić is postgraduate student at Faculty of Science, University of Sarajevo (Department of biology) with a special interest in field of molecular biology and bioinformatics. In diploma thesis has analyzed a codon usage in selected representatives of family Rhinolophidae (Chiroptera) with a dedication to implement different software packages towards improving the quality of his investigations.

Lada Lukić Bilela is associated professor of Molecular biology and Genomics at Faculty of Science, University of Sarajevo (Department of biology). She received her BSc in Biology (Faculty of Science, Sarajevo) MS and PhD in Mitochondrial genomics of Porifera at (Faculty of Science, Zagreb; field: molecular and cell biology). She was employed in Ruđer Bošković Institute at Laboratory of molecular genetics and performed her postdoctoral training at Johannes Gutenberg University of Mainz (Marie Curries Training Network fellowship).