Evidence for Compensatory Evolution of Ribosomal Proteins in Response to Rapid Divergence of Mitochondrial rRNA

Felipe S. Barreto*[,1] and Ronald S. Burton[1]

[1]Marine Biology Research Division, Scripps Institution of Oceanography, University of California, San Diego

***Corresponding author:** E-mail: fbarreto@ucsd.edu.

Associate editor: Willie Swanson

Abstract

Rapid evolution of mitochondrial DNA (mtDNA) places intrinsic selective pressures on many nuclear genes involved in mitochondrial functions. Mitochondrial ribosomes, for example, are composed of mtDNA-encoded ribosomal RNAs (rRNAs) and a set of more than 60 nuclear-encoded ribosomal proteins (mRP) distinct from the cytosolic RPs (cRP). We hypothesized that the rapid divergence of mt-rRNA would result in rapid evolution of mRPs relative to cRPs, which respond to slowly evolving nuclear-encoded rRNA. In comparisons of rates of nonsynonymous and synonymous substitutions between a pair of divergent populations of the copepod *Tigriopus californicus*, we found that mRPs showed elevated levels of amino acid changes relative to cRPs. This pattern was equally strong at the interspecific level, between three pairs of sister species (*Nasonia vitripennis* vs. *N. longicornis*, *Drosophila melanogaster* vs. *D. simulans*, and *Saccharomyces cerevisae* vs. *S. paradoxus*). This high rate of mRP evolution may result in intergenomic incompatibilities between taxonomic lineages, and such incompatibilities could lead to dysfunction of mitochondrial ribosomes and the loss of fitness observed among interpopulation hybrids in *T. californicus* and interspecific hybrids in other species.

Key words: rRNA, *Tigriopus*, ribosomal proteins, mitonuclear coadaptation.

Letter

In the course of speciation, complete reproductive isolation between populations is often preceded by a stage of reduced fitness of interpopulation hybrids. According to the widely discussed Dobzhansky–Muller model, such hybrid breakdown results from interlocus incompatibilities that arise in allopatry; alleles within populations are coadapted, and hybridization yields mismatched genotypes where coadaptation has been disrupted. Identification of loci involved in these incompatibilities is a key step in understanding the molecular mechanisms of postzygotic isolation (Coyne and Orr 2004).

Numerous interspecific genetic incompatibilities shown to cause hybrid inviability or infertility, mostly in *Drosophila*, involve interactions between nuclear-encoded gene products (Orr 2005; Brideau et al. 2006). Alternatively, hybrid breakdown can result from the disruption of interactions between nuclear and mitochondrial genomes (Rand et al. 2004; Burton et al. 2006; Burton and Barreto 2012). Mitonuclear interactions are particularly attractive as candidate genetic systems because 1) all 13 mitochondrial DNA (mtDNA)-encoded proteins and 24 RNAs (2 rRNAs and 22 tRNAs) require intimate interaction with nuclear gene products to achieve biological function, 2) their central role in cellular function means small perturbations of mitochondrial activity may have large impacts on fitness, and 3) the generally higher mutation rates in mtDNA compared with the nuclear genome provide a potentially strong selective pressure for compensatory changes in the interacting nuclear-encoded genes.

The intertidal copepod *Tigriopus californicus* has served as a model for testing predictions of mitonuclear incompatibility. Despite showing mtDNA divergence as high as 18% (Burton and Lee 1994; Edmands 2001), interpopulation crosses in the laboratory produce F_2 hybrids that are viable and fertile, but that experience significant breakdown in life history traits (Burton 1990; Edmands 1999) and mitochondrial function (Edmands and Burton 1999; Ellison and Burton 2006). Similar patterns of high mtDNA divergence and F_2 hybrid breakdown have also been documented in interspecific hybrids of *Nasonia* wasps (Breeuwer and Werren 1995; Oliveira et al. 2008), with mitonuclear incompatibilities implicated as major causes of postzygotic barriers (Niehuis et al. 2008). Although enzyme complexes of the oxidative phosphorylation system have received the most attention in studies of intergenomic coadaptation in these and other systems (Rand et al. 2004; Burton et al. 2006; Ellison et al. 2008; Lee et al. 2008), other epistatic mitonuclear complexes should not be overlooked (Burton and Barreto 2012).

Ribosomal proteins (RPs) associate closely with ribosomal RNA (rRNA) in the formation of ribosomes required for gene translation. A growing catalog of human disorders has been linked to ribosome dysfunction caused by mutations in RPs (Scheper et al. 2007). Although all RPs in eukaryotes (~160 total) are encoded by the nuclear genome, approximately half of them are imported and function exclusively in the mitochondria, forming ribosomes with mtDNA-encoded rRNAs. Notably, animal and fungal mitochondrial rRNAs evolve at faster rates than the nuclear rRNAs (table 1); consequently, we hypothesize that selection for compensatory mutations will be higher for RPs acting in

Table 1. Percent Divergence of rRNA Sequences in Pairs of Closely Related Taxa.

Species Pair	Nuclear-Encoded rRNA		mtDNA-Encoded rRNA	
	Small Subunit	Large Subunit	Small Subunit	Large Subunit
Tigriopus californicus SD–*T. californicus* SC[a]	0.68	0.20	7.5	13.1
Nasonia vitripennis–N. longicornis	0.29	0.94	7.4	6.9
Drosophila melanogaster–D. simulans	0.92	0.80	2.2	1.9
Saccharomyces cerevisae–S. paradoxus	0.38	0.01	1.9	2.4

[a]Comparison in *T. californicus* is between San Diego (SD) and Santa Cruz (SC) populations.

the mitochondria (mRPs) than for RPs acting in the cytoplasm (cRPs).

We took advantage of recently sequenced transcriptomes from two populations of *T. californicus* (Barreto et al. 2011) to test whether mRPs show higher rates of functional evolution than cRPs, as predicted by the mitonuclear coadaptation hypothesis. After identifying 65 mRPs and 70 cRPs and their correct open reading frames in both San Diego and Santa Cruz transcriptomes, we computed the rates of synonymous (d_S) and nonsynonymous (d_N) substitutions between the two populations for each gene. We found that mRPs and cRPs showed no differences in d_S (Mann–Whitney *U* test, $P = 0.91$). However, both d_N and d_N/d_S were significantly larger in mRPs compared with cRPs ($P < 1 \times 10^{-11}$ in both comparisons; fig. 1*A*). Among cRP genes, 49 (70%) had zero amino acid replacements, while only 2 (3%) mRPs had no changes.

Nucleotide diversity and substitution rates in coding DNA may be influenced by protein hydropathicity (Lobry and Gautier 1994), third position GC content (GC3; Moriyama and Gojobori 1992; Williams and Hurst 2000), and gene expression levels (Subramanian and Kumar 2004). Hence, we assessed whether these factors could explain the large difference in d_N/d_S between cRPs and mRPs. Although the two protein groups did not differ in hydropathicity ($F_{1,132} = 1.15$, $P = 0.28$), cRP genes showed significantly higher transcript levels ($F_{1,132} = 390$, $P < 0.0001$) and GC3 ($F_{1,132} = 121$, $P < 0.0001$) than mRP genes. The association of higher expression level with lower functional change seen in cRP genes is consistent with previous genomic studies (Pál et al. 2001; Subramanian and Kumar 2004). We found, however, that expression level is not correlated with d_N/d_S within each gene group (cRP: Spearman's $\rho = -0.046$, $P = 0.71$; mRP: Spearman's $\rho = -0.124$, $P = 0.32$). In addition, the d_N/d_S difference between cRPs and mRPs remained highly significant after accounting for variation in expression (analysis of covariance [ANCOVA]: $F_{1,131} = 23.3$, $P < 0.00001$).

Selection for biased codon usage may influence synonymous substitution rates (Ikemura 1985; Akashi 1994). Such bias is often strongly associated with GC3 content, but the direction and strength of the association may vary across taxa and functional gene groups (reviewed in Duret 2002). We detected no significant correlation between GC3 and d_S across RP genes in *T. californicus* (Spearman's $\rho = 0.054$, $P = 0.53$) and continued to observe highly significant differences in d_N/d_S between gene groups when accounting for GC3 variation (ANCOVA: $F_{1,131} = 53.5$, $P < 0.00001$). Taken together, the above results suggest the observed difference in functional divergence between cRPs and mRPs is, at least partly, independent of hydropathicity, GC3 content, and expression levels. Results from an analysis using information-theoretic approaches (Akaike's information criterion) also suggest that protein class (mRP or cRP) better explains d_N/d_S variation than do GC3 and expression (or combinations of the three factors) (supplementary methods and table S1, Supplementary Material online).

We tested the generality of this functional divergence pattern in three additional systems used as models in studies of Dobzhansky–Muller hybrid incompatibilities, all of which have higher divergence in mtDNA-encoded than in nuclear-encoded rRNAs (table 1). We retrieved and analyzed nearly complete sets of RP genes from *Nasonia vitripennis/longicornis* (sample sizes: cRP – 80, mRP – 65), *Drosophila melanogaster/simulans* (cRP – 73, mRP – 72), and *Saccharomyces cerevisae/paradoxus* (cRP – 78, mRP – 40), as described earlier. In agreement with the hypothesis of compensatory evolution, d_N and the d_N/d_S ratio for mRPs were significantly higher than those for cRPs in all comparisons ($P < 1 \times 10^{-7}$; fig. 1*B–D*), despite moderately higher d_S for mRPs ($P < 1 \times 10^{-5}$). As in *Tigriopus*, only a small proportion of mRPs (4.4–7.5%) had no amino acid replacements, while most cRPs (57–86%) were identical between sister species. Although we did not examine the influence of GC3 and gene expression in these latter data sets, the higher denominator (d_S) values found in mRPs (fig. 1*B–D*) suggest that, if a bias exists, it is likely artificially decreasing d_N/d_S for that set of genes (Wolf et al. 2009). Therefore, the difference in d_N/d_S between the groups is due to an extreme increase in d_N for mRPs rather than a difference in d_S.

Our results are consistent with a pattern of evolution in which the accumulation of mildly deleterious mutations in the rapidly evolving mitochondrial rRNA creates intrinsic selection pressure favoring compensatory mutations on the interacting RPs. As a result, amino acid replacements occur at a faster rate in mitochondrial-targeted RPs than in their cytosolic counterparts, even though both types are encoded in the nuclear genome. Although cRPs are clearly under strong purifying selection ($d_N/d_S \approx 0$), most mRPs showed d_N/d_S between zero and one, with only two cases (both in *Nasonia*) with $d_N/d_S > 1$. Such d_N/d_S ratios found for mRPs can result from either positive selection or relaxed functional constraints. When calculated across the entire gene sequence, however, as in this study, a criterion of $d_N/d_S > 1$ as indicative of positive selection is extremely stringent (Crandall et al. 1999; Swanson et al. 2001), since only a few codons are

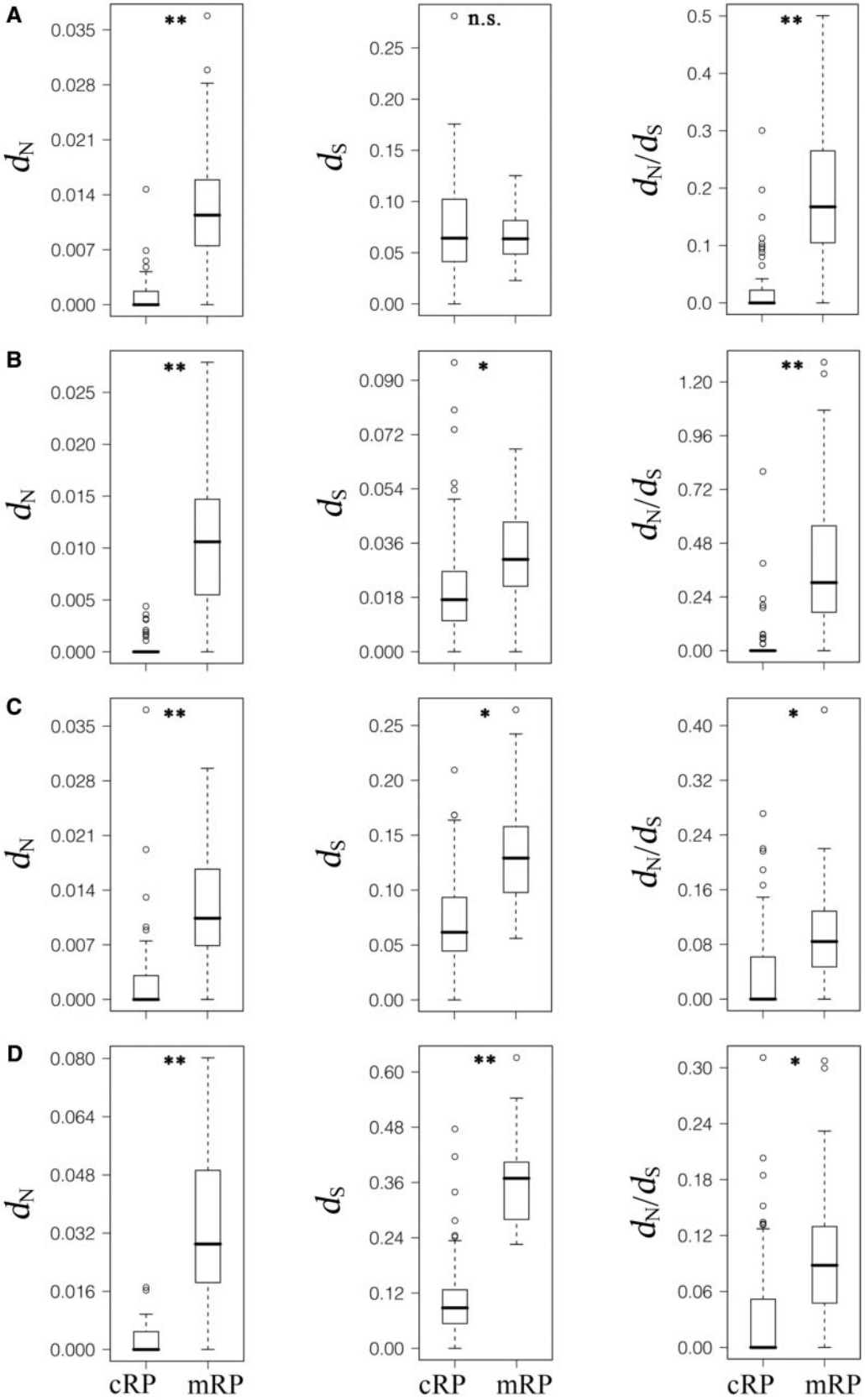

FIG. 1. Boxplot distributions of the nonsynonymous substitution rate (d_N), the synonymous substitution rate (d_S), and their ratio (d_N/d_S) for ribosomal protein (RP) genes that function in the cytoplasm (cRP) and in the mitochondria (mRP). Rates were estimated from pairwise alignments between (*A*) two populations of *Tigriopus californicus*, (*B*) *Nasonia vitripennis* and *N. longicornis*, (*C*) *Drosophila melanogaster* and *D. simulans*, and (*D*) *Saccharomyces cerevisae* and *S. paradoxus*. Statistical significance: $^{**}P < 10^{-15}$, $^{*}P < 10^{-7}$, n.s., not significant.

expected to be under direct positive selection (Clark et al. 2007). Swanson et al. (2004) showed that, upon more detailed analyses, evidence for positive selection (i.e., at least one codon with $d_N/d_S > 1$) was found in more than 30% of genes that showed overall d_N/d_S ratios between 0 and 0.2. It is hence likely that at least some mRPs in the systems investigated in our study have diverged due to positive selection, and future studies should focus on increasing taxon sampling in genes of interest and using statistical approaches that detect heterogenous d_N/d_S ratios across codons in each gene (Nielsen and Yang 1998; Yang et al. 2000).

Differences in the evolutionary rates of mRPs versus cRPs were first observed in electrophoretic comparisons of these proteins between highly divergent mammals (bovine to rat) (Matthews et al. 1978), while Werren et al. (2010) detected a significant enrichment of structural components of mitoribosomes in *Nasonia* within groups of genes with elevated d_N/d_S ratios. Our study took advantage of recent genome-level sequence information to demonstrate that, after correction for neutral mutation rate variation, mRP genes consistently showed higher levels of amino acid replacements than cRP genes across a range of taxa, which is suggestive of compensatory coadaptation (Rand et al. 2004). Moreover, our study shows that mRPs evolve rapidly even at the intraspecific level, providing a potential source of genetic incompatibility at the early stages of speciation.

Materials and Methods

Protein and cDNA sequences of the complete suite of cRP from *D. melanogaster* and *S. cerevisae*, as well as that of mRP from *D. melanogaster*, were downloaded from the Ribosomal Protein Gene Database (http://ribosome.miyazaki-med.ac.jp/, cited 2012 Apr 10; Nakao et al. 2004). The set of mRP genes from the yeast were downloaded from GenBank (NCBI) following Graack and Wittmann-Liebold (1998). Both sets of genes from *D. simulans* and *N. vitripennis* were obtained by BlastX searches of their genome databases in NCBI using the downloaded *D. melanogaster* sequences as queries. Since the genomes of *N. longicornis* and *S. paradoxus* are not fully annotated in public databases, we downloaded the most recent whole genome shotgun sequences of these species and formatted them as searchable databases using NCBI's standalone scripts. Their cRP and mRP gene sequences were then identified through BlastN searches using their respective congener's sequences as queries. Exons were retrieved and joined using custom Perl scripts.

Transcriptomes from two populations of *T. californicus* were assembled for a previous study (Barreto et al. 2011). Even though the contigs had been previously annotated, we employed an additional bioinformatic procedure to confirm and fine-tune the annotations with regard to RP identity. Using NCBI's standalone scripts, we performed reciprocal Blast searches between each of our *T. californicus* transcriptomes and the *D. melanogaster* RPs, using BlastX and TBlastN accordingly. Only sequences that were each other's reciprocal best hits were retained for further analyses. Finally, to reduce the chances of including sequencing errors in our divergence estimates, we retained only contigs that had mean coverage $\geq$ ten reads/base pair when they were assembled (Barreto et al. 2011).

For sequences identified through Blast searches mentioned earlier, the most likely open reading frame for each RP gene was extracted based on the best BlastX hit. Each pair of orthologous RP genes between sister taxa was aligned in ClustalW, and all alignments were visually inspected. Only alignments longer than 100 bp were kept, and d_N, d_S, and d_N/d_S were estimated in PAML (Yang 2007) using the maximum likelihood method implemented in the package YN00 (Yang and Nielsen 2000). We employed a Mann–Whitney *U* test of the null hypothesis that d_N, d_S, and d_N/d_S did not differ between mRPs and cRPs.

Alignments of rRNA regions were also performed in ClustalW, and the percentage of polymorphic sites was assessed in MEGA5 (Tamura et al. 2011). The coding regions for the *T. californicus* RPs identified in this study have been deposited in GenBank (accession numbers are found in supplementary table S2, Supplementary Material online).

For the RP coding sequences in *T. californicus*, we calculated GC3 as well as protein hydropathicity (Kyte and Doolittle 1982) using the program CodonW (http://codonw.sourceforge.net, cited 2012 Aug 14). Expression of these genes was quantified by mapping the original 454 reads from Barreto et al. (2011) onto the contigs using the program CLC Genomics Workbench 5.1 (CLC Bio) and then normalizing read counts by contig length and total mapped reads. For each gene, we averaged the values for GC3, hydropathicity, and $\log_2$ gene expression between populations and then tested whether these parameters differed between cRPs and mRPs by means of analysis of variances. When a difference was detected, we assessed the role of the respective variable on substitution rates using Spearman's rank correlations and ANCOVA.

Supplementary Material

Supplementary methods and tables S1 and S2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank C.K. Ellison and R. Pereira for insightful discussions and are also grateful to Associate Editor Willie Swanson and an anonymous reviewer for constructive suggestions on the manuscript. This work was supported by the National Science Foundation (grant no. DEB1051057 to R.S.B). Sequences for *T. californicus* used in this study were deposited in GenBank. Accession numbers used in this study can be found in the supplementary material, Supplementary Material online.

References

Akashi H. 1994. Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accurancy. *Genetics* 136:927–935.

Barreto FS, Moy GW, Burton RS. 2011. Interpopulation patterns of divergence and selection across the transcriptome of the copepod *Tigriopus californicus*. *Mol Ecol.* 20:560–572.

Breeuwer JAJ, Werren JH. 1995. Hybrid breakdown between two haplodiploid species: the role of nuclear and cytoplasmic genes. *Evolution* 49:705–717.

Brideau NJ, Flores HA, Wang J, Maheshwari S, Wang X, Barbash DA. 2006. Two Dobzhansky–Muller genes interact to cause hybrid lethality in *Drosophila*. *Science* 314:1292–1295.

Burton RS. 1990. Hybrid breakdown in developmental time in the copepod *Tigriopus californicus*. *Evolution* 44:1814–1822.

Burton RS, Barreto FS. 2012. A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? *Mol Ecol.* 21(20): 4942–4957.

Burton RS, Ellison CK, Harrison JS. 2006. The sorry state of F2 hybrids: consequences of rapid mitochondrial DNA evolution in allopatric populations. *Am Nat.* 168:S14–S24.

Burton RS, Lee BN. 1994. Nuclear and mitochondrial gene genealogies and allozyme polymorphism across a major phylogeographic break in the copepod *Tigriopus californicus*. *Proc Natl Acad Sci U S A.* 91: 5197–5201.

Clark AG, Eisen MB, Smith D, et al. (14 co-authors). 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450: 203–218.

Coyne JA, Orr HA. 2004. Speciation. Sunderland (MA): Sinauer Associates.

Crandall KA, Kelsey CR, Imamichi H, Lane HC, Salzman NP. 1999. Parallel evolution of drug resistance in HIV: failure of nonsynonymous/synonymous substitution rate ratio to detect selection. *Mol Biol Evol.* 16: 372–382.

Duret L. 2002. Evolution of synonymous codon usage in metazoans. *Curr Opin Genet Dev.* 12:640–649.

Edmands S. 1999. Heterosis and outbreeding depression in interpopulation crosses spanning a wide range of divergence. *Evolution* 53: 1757–1768.

Edmands S. 2001. Phylogeography of the intertidal copepod *Tigriopus californicus* reveals substantially reduced population differentiation at northern latitudes. *Mol Ecol.* 10:1743–1750.

Edmands S, Burton RS. 1999. Cytochrome c oxidase activity in interpopulation hybrids of a marine copepod: a test for nuclear–nuclear or nuclear–cytoplasmic coadaptation. *Evolution* 53:1972–1978.

Ellison CK, Burton RS. 2006. Disruption of mitochondrial function in interpopulation hybrids of *Tigriopus californicus*. *Evolution* 60: 1382–1391.

Ellison CK, Niehuis O, Gadau J. 2008. Hybrid breakdown and mitochondrial dysfunction in hybrids of *Nasonia* parasitoid wasps. *J Evol Biol.* 21:1844–1851.

Graack HR, Wittmann-Liebold B. 1998. Mitochondrial ribosomal proteins (MRPs) of yeast. *Biochem J.* 329:433–448.

Ikemura T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol.* 2:13–34.

Kyte J, Doolittle R. 1982. A simple method for displaying the hydropathic character of a protein. *J Mol Biol.* 157:105–132.

Lee H-Y, Chou J-Y, Cheong L, Chang N-H, Yang S-Y, Leu J-Y. 2008. Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell* 135:1065–1073.

Lobry JR, 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: 3174–3180.

Matthews DE, Hessler RA, O'Brien TW. 1978. Rapid evolutionary divergence of proteins in mammalian mitochondrial ribosomes. *FEBS Lett.* 86:76–80.

Moriyama EN, Gojobori T. 1992. Rates of synonymous substitutions and base composition of nuclear genes in *Drosophila*. *Genetics* 130:855–864.

Nakao A, Yoshihama M, Kenmochi N. 2004. RPG: the Ribosomal Protein Gene database. *Nucleic Acids Res.* 32:D168–D170.

Niehuis O, Judson AK, Gadau J. 2008. Cytonuclear genic incompatibilities cause increased mortality in male F_2 hybrids of *Nasonia girault*i and *N. vitripennis*. *Genetics* 178:413–426.

Nielsen R, Yang Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148:929–936.

Oliveira DCSG, Raychoudhury R, Lavrov DV, Werren JH. 2008. Rapidly evolving mitochondrial genome and directional selection in mitochondrial genes in the parasitic wasp *Nasonia* (Hymenoptera: Pteromalidae). *Mol Biol Evol.* 25:2167–2180.

Orr HA. 2005. The genetic basis of reproductive isolation: insights from *Drosophila*. *Proc Natl Acad Sci U S A.* 102:6522–6526.

Pál C, Papp B, Hurst LD. 2001. Highly expressed genes in yeast evolve slowly. *Genetics* 158:927–931.

Rand DM, Haney RA, Fry AJ. 2004. Cytonuclear coevolution: the genomics of cooperation. *Trends Ecol Evol.* 19:645–653.

Scheper GC, van der Knaap MS, Proud CG. 2007. Translation matters: protein synthesis defects in inherited disease. *Nat Rev Genet.* 8: 711–723.

Subramanian S, Kumar S. 2004. Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome. *Genetics* 168:373–381.

Swanson WJ, Clark AG, Waldrip-Dail H, Wolfner MF, Aquadro CF. 2001. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *Proc Natl Acad Sci U S A.* 98:7375–7379.

Swanson WJ, Wong A, Wolfner MF, Aquadro CF. 2004. Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive selection. *Genetics* 168: 1457–1465.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 28:2731–2739.

Werren JH, Richards S, Desjardins CA, et al. (169 co-authors). 2010. Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science* 327:343–348.

Williams EJB, Hurst LD. 2000. The proteins of linked genes evolve at similar rates. *Nature* 407:900–903.

Wolf JBW, Künstner A, Nam K, Jakobsson M, Ellegren H. 2009. Nonlinear dynamics of nonsynonymous (d_N) and synonymous (d_S) substitution rates affects inference of selection. *Genome Biol Evol.* 1:308–319.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.

Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol.* 17:32–43.

Yang Z, Nielsen R, Goldman N, Pedersen AK. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431–449.