The Roles of Mutation, Selection, and Expression in Determining Relative Rates of Evolution in Mitochondrial versus Nuclear Genomes

Justin C. Havird* and Daniel B. Sloan*
Department of Biology, Colorado State University, Fort Collins, CO
*Corresponding authors: E-mails: justin.havird@colostate.edu; dbsloan@rams.colostate.edu.
Associate editor: David Irwin

Abstract

Eukaryotes rely on proteins encoded by the nuclear and mitochondrial (mt) genomes, which interact within multisubunit complexes such as oxidative-phosphorylation enzymes. Although selection is thought to be less efficient on the asexual mt genome, in bilaterian animals the ratio of nonsynonymous to synonymous substitutions (ω) is lower in mt- compared with nuclear-encoded OXPHOS subunits, suggesting *stronger* effects of purifying selection in the mt genome. Because high levels of gene expression constrain protein sequence evolution, one proposed resolution to this paradox is that mt genes are expressed more highly than nuclear genes. To test this hypothesis, we investigated expression and sequence evolution of mt and nuclear genes from 84 diverse eukaryotes that vary in mt gene content and mutation rate. We found that the relationship between mt and nuclear ω values varied dramatically across eukaryotes. In contrast, transcript abundance is consistently higher for mt genes than nuclear genes, regardless of which genes happen to be in the mt genome. Consequently, expression levels cannot be responsible for the differences in ω. Rather, 84% of the variance in the ratio of ω values between mt and nuclear genes could be explained by differences in mutation rate between the two genomes. We relate these findings to the hypothesis that high rates of mt mutation select for compensatory changes in the nuclear genome. We also propose an explanation for why mt transcripts consistently outnumber their nuclear counterparts, with implications for mitonuclear protein imbalance and aging.

Key words: **mitonuclear interactions, cytonuclear interactions, transcription, d_N/d_S, positive selection, nuclear compensation.**

Introduction

In order to generate the ATP necessary for basic biological functions, eukaryotes rely on gene products encoded by both the nuclear genome and a much smaller mitochondrial (mt) genome. Specifically, a mixture of nuclear- and mt-encoded products physically interact to form the multisubunit oxidative phosphorylation (OXPHOS) enzyme complexes (Allen 2015) and components of the mt translational machinery (particularly ribosomes; Maier et al. 2013). The nuclear and mt genomes are replicated and inherited in fundamentally different ways, resulting in fascinating coevolutionary dynamics that have important consequences for health, aging, and eukaryotic diversification (Dowling 2014; Hill 2015). In most animals, the rate of sequence evolution in the mt genome is much faster than in the nuclear genome, reflecting a higher underlying mutation rate (Brown et al. 1979). In addition, most mt genomes are uniparentally inherited and effectively haploid, which is expected to reduce their effective population size and the efficiency of selection against deleterious mutations (Lynch and Blanchard 1998; Neiman and Taylor 2009). Therefore, mildly deleterious mutations should accumulate more rapidly in mt-encoded OXPHOS subunits relative to nuclear-encoded OXPHOS subunits. Surprisingly, however, mt-encoded OXPHOS subunits in animals tend to have a lower ratio of nonsynonymous to synonymous substitutions (d_N/d_S or ω) than nuclear-encoded OXPHOS subunits (Nabholz et al. 2013; Popadin et al. 2013), which is typically a signature of *more* efficient selection against harmful mutations (Nielsen 2005).

One explanation for this apparent paradox is that higher ω values in nuclear genes encoding mt-targeted proteins (N-mt genes) are not due to relaxed selection, but rather are the result of positive selection for compensatory amino-acid substitutions in response to rapidly accumulating mt mutations. We refer to this as the nuclear compensation hypothesis, which has been supported by the following: (1) patterns of mitonuclear evolution such as the position and relative timing of mt and N-mt substitutions at interacting sites within protein complexes, (2) the reduced fitness of cytoplasmic hybrids, (3) the preferential retention of nuclear loci from the maternal parent following allopolyploidy events, and (4) the preferential co-introgression of N-mt genes with mt genomes (Burton et al. 2006, 2013; Dowling et al. 2007; Osada and Akashi 2012; Meiklejohn et al. 2013; Gong et al. 2014; Beck et al. 2015; Havird et al. 2015).

Alternatively, N-mt genes may be less functionally important and therefore under less intense purifying selection than mt genes, resulting in higher ω values in N-mt genes (Nabholz

et al. 2013; Popadin et al. 2013; Zhang and Broughton 2013). We refer to this argument as the selective constraints hypothesis, which is supported by two main lines of reasoning. First, N-mt genes often encode "peripheral" subunits within OXPHOS complexes, which may be able to withstand more mutations due to a less critical role in OXPHOS (note that some of these peripheral subunits are eukaryotic-specific additions to these complexes and were not ancestrally present in bacteria; van der Sluis et al. 2015), whereas mt genes encode "core" subunits, which may participate in more essential functions (Zhang and Broughton 2013). Secondly, proteins that are highly abundant may be under more functional constraints because protein misfolding and misinteractions can be more costly for abundant proteins, resulting in stronger purifying selection and slower evolution (Zhang and Yang 2015). In studies supporting this relationship, transcript abundance is often used as a proxy for protein abundance. Accordingly, it has been shown that transcript abundance in animals is much higher (~20-fold) for mt OXPHOS subunits than N-mt ones, and expression can describe ~75% of the variance in ω among OXPHOS genes (Nabholz et al. 2013), suggesting that mt genes may be under stronger functional and selective constraints because of their high expression levels. However, there are important confounding factors in such analyses because essentially all bilaterian animals have mt genomes with the same set of 13 protein coding genes (all OXPHOS subunits) and mutation rates that are consistently higher than in the nucleus (Brown et al. 1979; Boore 1999; Smith and Keeling 2015). Therefore, it is difficult to attribute the difference in ω values between animal mt and nuclear OXPHOS genes to expression rather than mutation rate. Moreover, it is unclear whether the high levels of expression observed for mt genes are a property of the mt genome per se or just the specific set of genes that happen to be found in animal mt genomes.

From a broader evolutionary perspective, mt gene content and rates of sequence evolution vary greatly across different eukaryotic lineages (Wolfe et al. 1987; Gray et al. 1999; Smith and Keeling 2015). For example, the large mt genome of the jakobid *Andalucia godoyi* contains 100 genes (Burger et al. 2013). In contrast, the tiny mt genome of the alveolate *Chromera velicia* appears to contain only a single gene (Petersen et al. 2014), and some parasitic and anaerobic eukaryotes have entirely lost their electron transport chain and mt genome (Hjort et al. 2010; Karnkowska et al. 2016). The variation in the number of mt genes across eukaryotes largely stems from the process of endosymbiotic gene loss and transfer to the nuclear genome that is ongoing in many lineages (Adams and Palmer 2003; Timmis et al. 2004). For example, mt gene loss/transfer is highly active in angiosperms, resulting in mt genomes that typically contain anywhere from 24 to 41 protein-coding genes (Adams et al. 2002; Mower and Bonen 2009). Nearly all angiosperm mt genomes contain a core gene set comprising 18 OXPHOS genes (including all 13 of the mt-encoded proteins found in most bilaterian animals), four genes involved in cytochrome C biogenesis, and two genes (*matR* and *mttB*) that may play a role in intron splicing and membrane transport, respectively (cf. Skippington et al. 2015). There are 17 more genes that were present in the common ancestor of angiosperms and have been retained to varying degrees in some extant lineages but lost in others. These include two genes (*sdh3* and *sdh4*) that encode subunits of OXPHOS complex II (succinate dehydrogenase) and 15 ribosomal protein genes (Adams et al. 2002; Mower and Bonen 2009). A particularly interesting example of ongoing intracellular gene transfer involves *cox2* (Adams et al. 1999). This OXPHOS gene is part of the core mt gene set in plants and the vast majority of other eukaryotes, but it was recently duplicated and transferred to the nucleus during the evolution of legumes. The ancestral mt copy of *cox2* has since been lost from the mt genome in some species but retained in others.

The mt genomes of angiosperms also fundamentally differ from those of bilaterian animals in that they have much lower rates of nucleotide substitution than the nuclear genome (Wolfe et al. 1987), which may be due to mt recombination/repair mechanisms found in plants but not animals (Christensen 2014). However, the slow-plant/fast-animal dichotomy of mt evolution is a gross oversimplification. Some nonbilaterian animals show low rates of mt evolution (Lavrov 2007), and several independent angiosperm lineages have dramatically increased mt substitution rates that are on par with those of many bilaterian animals (Mower et al. 2007). Interestingly, early comparisons between rates of mt and nuclear sequence evolution in multiple eukaryotic lineages provided hints that the relationship between mt and nuclear ω values that has been observed in animals may differ substantially in other evolutionary groups (Lynch and Blanchard 1998).

Here, we analyze expression levels and evolutionary rates of mt and N-mt genes across diverse eukaryotic lineages. We mainly focus on genes encoding OXPHOS subunits, as these tend to be preferentially retained in mitochondrial genomes across eukaryotes (Allen 2015). We show that high levels of gene expression are a ubiquitous feature of mt genomes, regardless of the specific genes that they encode. However, differences in ω between mt and N-mt genes are highly variable among eukaryotes and can largely be explained by differences in neutral substitution rates rather than gene expression. We discuss the implications of these results for the nuclear compensation and selective constraints hypotheses. In addition, we propose functional explanations for the surprising finding that mt transcripts drastically outnumber N-mt transcripts, even though mt and N-mt subunits generally exist at equimolar ratios within OXPHOS complexes.

Results and Discussion

Signatures of Selection on Mitochondrial- and Nuclear-Encoded OXPHOS Subunits Vary Widely across Eukaryotes

In animals, ω values for mt OXPHOS subunits (ω_{mt}) are much lower than for N-mt OXPHOS subunits (ω_{nuc}) (fig. 1; Nabholz et al. 2013; Popadin et al. 2013). However, we found that this pattern varied dramatically among eukaryotic lineages (table 1), as the ω_{mt}/ω_{nuc} ratio ranged from 0.19 in carnivores to >2 in many angiosperms (fig. 1; supplementary

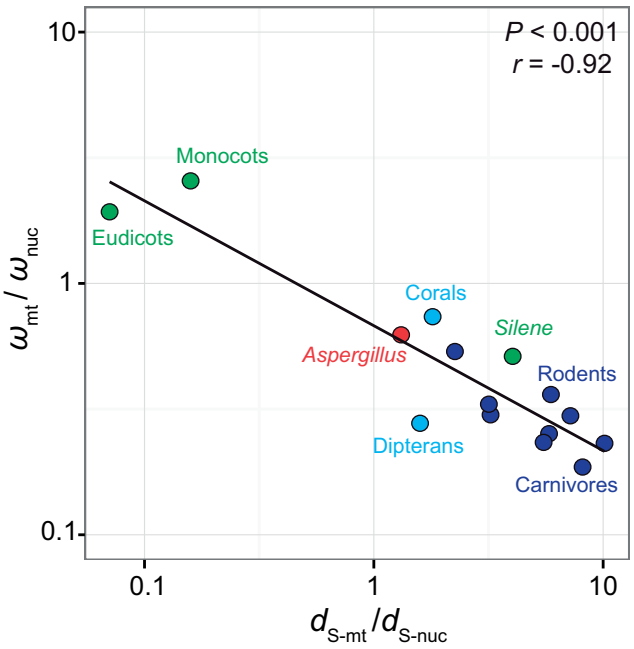

Fig. 1. Signatures of selection on mt- and nuclear-encoded OXPHOS genes across eukaryotic lineages. The ratio of mt to nuclear synonymous substitution rates ($d_{S\text{-}mt}/d_{S\text{-}nuc}$) largely explains the variance in ω_{mt}/ω_{nuc} observed across eukaryotic lineages (vertebrates are in dark blue, invertebrates in light blue, angiosperms in green, and fungi in red). In addition to rodents and carnivores, vertebrate lineages included primates, bats, cetartiodactyls, birds, reptiles, amphibians, and teleosts. The species and data used are shown in supplementary table S2 and supplementary data file S2, Supplementary Material online, respectively.

figs. S1 and S2, Supplementary Material online). To investigate the role of mutation in this variation, we used synonymous substitution rates ($d_{S\text{-}mt}$ and $d_{S\text{-}nuc}$ for the mt and nuclear genomes, respectively) to approximate underlying mutation rates (with the caveat that phylogenetic comparisons may underestimate the true mutation rate because of selection on synonymous changes and other factors; Denver et al. 2000; Chamary et al. 2006; Lynch et al. 2008). We found that there was a striking negative relationship between ω_{mt}/ω_{nuc} and $d_{S\text{-}mt}/d_{S\text{-}nuc}$ ($P < 0.001$, $r = -0.92$; fig. 1). In other words, species with faster-evolving mtDNA had lower mt ω values relative to the nucleus. Importantly, this relationship remained strong after controlling for phylogeny ($P < 0.001$, $r = -0.92$; supplementary fig. S3, Supplementary Material online), as illustrated by the fact that angiosperms with unusually fast-evolving mt genomes (from the genus *Silene*; Sloan et al. 2009) grouped with bilaterian animals rather than with other angiosperms. Because ω is thought to reflect the mixture of positive and purifying selection acting on amino acid substitutions, the strong relationship between ω_{mt}/ω_{nuc} and $d_{S\text{-}mt}/d_{S\text{-}nuc}$ across eukaryotes (fig. 1) suggests that relative selection pressures acting on the mt and nuclear genomes systematically vary with relative rates of mt mutation.

The Nuclear Compensation Hypothesis

The negative relationship between ω_{mt}/ω_{nuc} and $d_{S\text{-}mt}/d_{S\text{-}nuc}$ (fig. 1) is consistent with the nuclear compensation hypothesis, at least at a coarse level. As relative mt mutation rate ($d_{S\text{-}mt}/d_{S\text{-}nuc}$) increases, the relative fixation probability of amino acid substitutions in the nucleus (ω_{nuc}/ω_{mt}) also increases, which could indicate that higher mt mutation rates generate more positive selection for compensatory changes in N-mt genes. If compensatory mitonuclear evolution truly is the major cause of the observed relationship in figure 1, we would not expect to find this pattern for nuclear genes that are targeted to other parts of the cell besides the mitochondria. Therefore, we performed a parallel set of analyses using nuclear genes that encode components of the glycolysis metabolic pathway, along with the same set of mt-encoded OXPHOS subunits (table 1). Glycolysis genes serve as an appropriate control to test the nuclear compensation hypothesis because, like OXPHOS genes, they play a central role in energy metabolism, but they are targeted to the cytosol and have no mt interactions. In this analysis, we used the same methods as for the OXPHOS subunits, with the exception that the nuclear-encoded genes were from the glycolysis metabolic pathway. We found that the glycolysis genes exhibited a negative relationship between ω_{mt}/ω_{nuc} and $d_{S\text{-}mt}/d_{S\text{-}nuc}$ that was almost as strong as observed with OXPHOS genes ($P = 0.003$, $r = -0.71$; fig. 2A), demonstrating that this relationship cannot be explained solely by patterns of nuclear compensation. In fact, most of the observed effect is driven by the very strong negative relationship between ω_{mt} and $d_{S\text{-}mt}/d_{S\text{-}nuc}$ regardless of whether OXPHOS or glycolysis genes are used to estimate $d_{S\text{-}nuc}$ ($P < 0.002$, $r < -0.75$ for both; fig. 2B).

However, a more specific prediction under the nuclear compensation hypothesis is that mitochondrial mutation rates should be positively correlated with ω_{nuc} for OXPHOS genes but not glycolysis genes. In other words, higher rates of mt genome evolution should select for increased frequency of compensatory amino acid substitutions in N-mt genes but not in other nuclear genes. Our data provided some support for this prediction, as there was a significant, positive relationship between $d_{S\text{-}mt}/d_{S\text{-}nuc}$ and ω_{nuc} for OXPHOS genes ($P = 0.028$, $r = 0.56$) but not for glycolysis genes ($P = 0.712$, $r = 0.10$; fig. 2C). Although this effect explains only a small proportion of the observed relationship in figure 1, it does suggest that selection for compensatory nuclear substitutions in response to mt mutation accumulation can have detectable effects on rates of sequence evolution, even across the enormous taxonomic breadth of our sampling. Future tests evaluating the nuclear compensation hypothesis that use finer-scale phylogenetic sampling and restrict molecular analyses to mitonuclear contact residues may be more powerful and help interpret the far-reaching implications of this hypothesis for ecology and evolution (Hill 2015).

Mitochondrial-Encoded Genes Are Consistently Expressed at Higher Levels than Their Nuclear Counterparts across Eukaryotes

One prevailing explanation for the observation that ω_{mt} values are much lower than ω_{nuc} values in animals is that mt genes are expressed at higher levels and thereby subject to more intense purifying selection (Nabholz et al. 2013;

Table 1. Summary Statistics for mt- and Nuclear-Encoded Genes from Eukaryotes Analyzed for ω Corresponding to Data in Figures 1 and 2.

Taxonomic group	Num taxa	Mt-OXPHOS subunits		Nuc-OXPHOS subunits		Nuc-glycolysis genes	
		Num genes	Length (bp)	Num genes	Length (bp)	Num genes	Length (bp)
Primates	9	13	10,902	70	54,489	28	44,664
Rodents	5	13	11,430	80	55,992	30	42,933
Cetartiodactyls	9	13	11,412	67	53,256	29	67,740
Carnivores	4	13	11,415	73	57,072	35	59,841
Bats	3	13	11,451	76	52,389	33	49,569
Birds	10	12	10,902	45	45,582	25	37,890
Reptiles	5	13	11,544	62	49,095	30	49,191
Amphibians	2	13	11,409	61	45,315	30	40,578
Teleosts	4	13	11,478	69	44,472	30	43,032
Dipterans	7	11	10,578	53	32,400	21	31,275
Corals	7	13	12,333	42	27,771	18	26,667
Silene (Eudicot)	3	18	14,208	48	26,163	18	21,669
Other eudicots	9	14	13,347	25	16,611	21	34,104
Monocots	5	15	9,624	30	21,051	20	32,007
Fungi (*Aspergillus*)	2	14	13,440	34	30,321	20	26,328
Total	84	201	175,473	835	611,979	388	607,488

Zhang and Yang 2015). This hypothesis yields the testable prediction that the relationship between levels of mt and N-mt gene expression should vary across eukaryotic lineages in conjunction with the observed variation in ω_{mt}/ω_{nuc}. However, we found that the difference in expression between mt and nuclear genomes is a seemingly ubiquitous feature of eukaryotes, even in lineages with relatively high ω_{mt} values (fig. 3). Using RNA-seq data, we first confirmed previous work in animals (fig. 3; Nabholz et al. 2013) and then determined that mt OXPHOS transcripts were also much more abundant than N-mt transcripts in plants (~18-fold; fig. 3) and in fungi (~6-fold; fig. 3). Of course, the vast majority of eukaryotic diversity is found in protist lineages. We did not include any protists in our analysis because RNA-seq in these systems is based almost exclusively on poly-A enrichment, which can bias estimates of mitochondrial transcript abundance (see Materials and Methods). Therefore, while our results are clearly applicable to diverse eukaryotes, they may not be universal.

Similar to the patterns for OXPHOS subunits, transcript abundance for mt-encoded ribosomal protein genes was higher than that of N-mt ribosomal protein genes in the model angiosperm *Arabidopsis* (~10-fold; supplementary fig. S4, Supplementary Material online). Moreover, plastid-encoded subunits involved in electron transport and protease activity were also expressed more highly than the corresponding nuclear-encoded subunits in angiosperms (~19-fold; fig. 3). These findings suggest that the difference in transcript abundance between nuclear and organelle genomes is a general phenomenon of cytonuclear protein complexes.

If the observed differences in transcript levels between mt and N-mt genes are a fundamental property of the genomes themselves, then when an mt gene is transferred to the nuclear genome, its transcript abundance should decrease to a typical nuclear level. We confirmed this prediction by taking advantage of the fact that intracellular gene transfer is ongoing in plants (Adams et al. 2002) and by sampling RNA-seq data sets from angiosperms with variable mt gene content (figs. 3 and 4). For example, in tobacco, the succinate dehydrogenase (SDH) genes *sdh3* and *sdh4* are mt-encoded and expressed at high levels relative to nuclear-encoded SDH subunits (average FPKM = 55,802 versus 1,357). In contrast, these genes have been transferred to the nuclear genome in *Arabidopsis* and show similar expression levels compared with other N-mt SDH subunits (average FPKM = 4,575 versus 4,780). We found a similar pattern for the transfer of the cytochrome c oxidase gene *cox2* to the nuclear genome in the legume *Vigna radiata* (fig. 4; Adams et al. 1999). Therefore, the variable nature of mt gene content in plants demonstrates that the observed differences in transcript abundance are fundamental properties of mt and nuclear genomes themselves rather than the specific genes that happen to be found in these genomes (e.g., core versus peripheral OXPHOS subunits).

Expression Differences Do Not Explain Contrasting Signatures of Selection in Mitochondrial and Nuclear Genes

We used RNA-seq data to calculate ω and quantify transcript abundance for individual mt and N-mt genes from angiosperms. In contrast to previous findings in bilaterian animals (fig. 5; Nabholz et al. 2013), we found that ω was generally not correlated with transcript abundance (fig. 5). However, a slight negative relationship similar to that in animals was observed for *Silene* species with accelerated mt substitution rates (Sloan et al. 2012), whereas a slight positive relationship was observed for angiosperms with slowly-evolving mt genomes (fig. 5). These results support the conclusion that differences in ω between nuclear and mt genomes are primarily associated with extreme variation in synonymous substitution rates across eukaryotes (fig. 1).

When we performed a similar analysis using plastid- and nuclear-encoded genes that form multisubunit complexes functioning in electron transport and photosynthesis in plastids, we also failed to observe a negative relationship between ω and transcript abundance (fig. 5 and supplementary fig. S5,

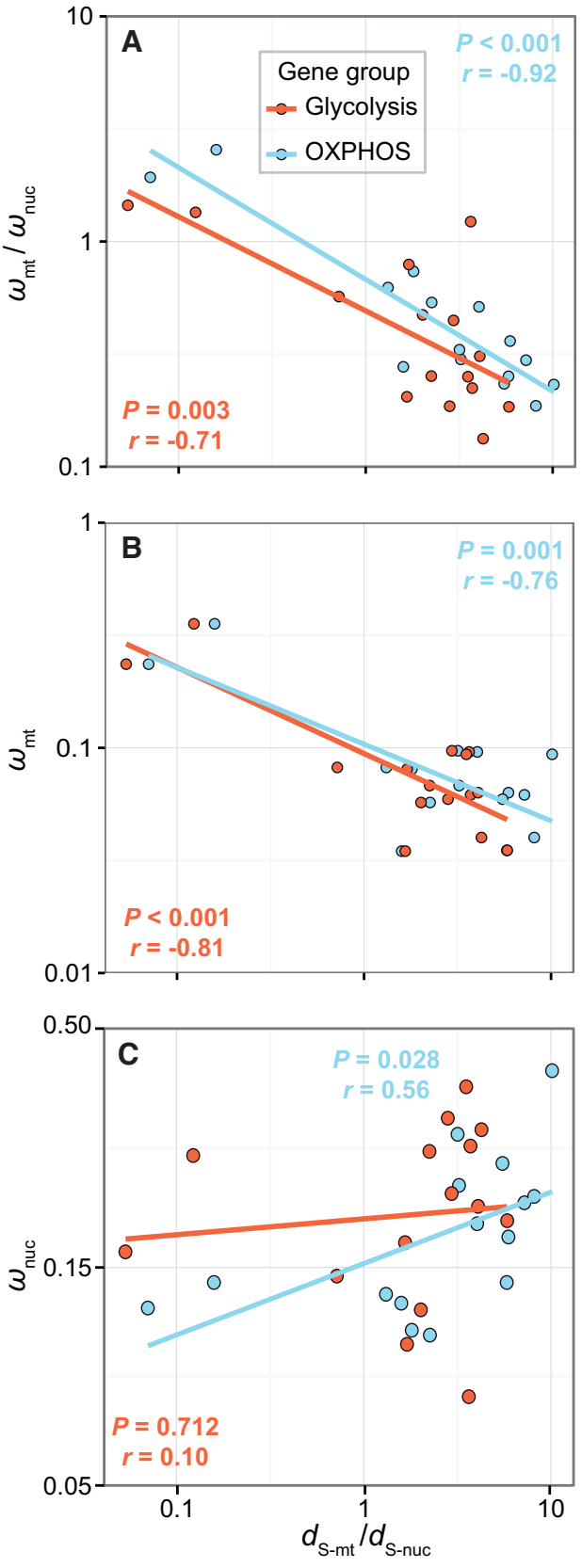

FIG. 2. Comparison between signatures of selection on mt and N-mt OXPHOS and glycolysis genes. The ratio of mt to nuclear synonymous substitution rates ($d_{S\text{-}mt}/d_{S\text{-}nuc}$) is significantly related to (A) ω_{mt}/ω_{nuc} and (B) ω_{mt} for both OXPHOS and glycolysis genes, but only to (C) ω_{nuc} for OXPHOS genes.

Supplementary Material online). Taken together, these results largely rule out one of the mechanisms (gene expression) that has been proposed under the selective constraints hypothesis. Likewise, our findings are inconsistent with arguments that the higher ω values for N-mt versus mt OXPHOS subunits in animals are primarily caused by reduced selective constraints on peripheral versus core proteins (Popadin et al. 2013; Zhang and Broughton 2013), because the opposite pattern is observed in plants (fig. 1). However, differences in selective constraints undoubtedly still shape the evolution of N-mt genes. For example, lower levels of gene expression and functional constraint are likely a major cause of the fast evolutionary rate in N-mt genes that encode mt ribosomal proteins and tRNA synthetases relative to the (non N-mt) nuclear genes that play analogous roles in the cytosol (Barreto and Burton 2013; Sloan et al. 2014; Adrion et al. 2015; Pett and Lavrov 2015). The former only translate transcripts from the small number of genes encoded in the mitochondria, whereas the latter are responsible for translating the much larger number of nuclear transcripts.

In light of the striking relationship described in figure 1, it is also important to scrutinize the biological meanings of ω and d_S. A common assumption when interpreting ω as a measure of selection is that it is independent from the mutation rate. This would imply that, all else being equal, as mutation rate increases, d_N and d_S both increase proportionally. However, ω is almost certainly not independent of mutation rate. When mutation rate is inferred from d_S, it is readily apparent that these two measures are statistically nonindependent because d_S is the denominator of the ω ratio. Genome-wide studies have shown that d_S and ω can be strongly correlated, and the correlation between ω and mutation rate may be similar in magnitude to the correlation between ω and selective strength (Wyckoff et al. 2005; Vallender and Lahn 2007). Moreover, other factors such as the substitution model, lineage-specific effects, and patterns of adjacent substitutions may influence the correlation between ω and the mutation rate in complex ways (Li et al. 2009; Wolf et al. 2009; Stoletzki and Eyre-Walker 2011). These complications, coupled with the fact that selection often acts on some synonymous substitutions (Chamary et al. 2006), make interpreting ω as a measure of selection more difficult than often assumed, especially when making direct comparisons between ω values from genomes with very different mutation rates (Nabholz et al. 2013; Popadin et al. 2013). As indicated by other recent analyses (Cooper et al. 2015), some classic conclusions about the strength and efficiency of selection in mt versus nuclear genomes may require reassessment.

Why Are Mitochondrial Transcripts More Abundant than Their Nuclear Counterparts?

Our results show that, across diverse eukaryotic lineages, genes encoded in the mt genome have higher transcript abundances than interacting genes that are encoded in the nucleus (figs. 3–5; supplementary figs. S4 and S5, Supplementary Material online; Nabholz et al. 2013). However, because subunits generally exist at equimolar ratios within OXPHOS complexes, it is unclear why mt transcripts

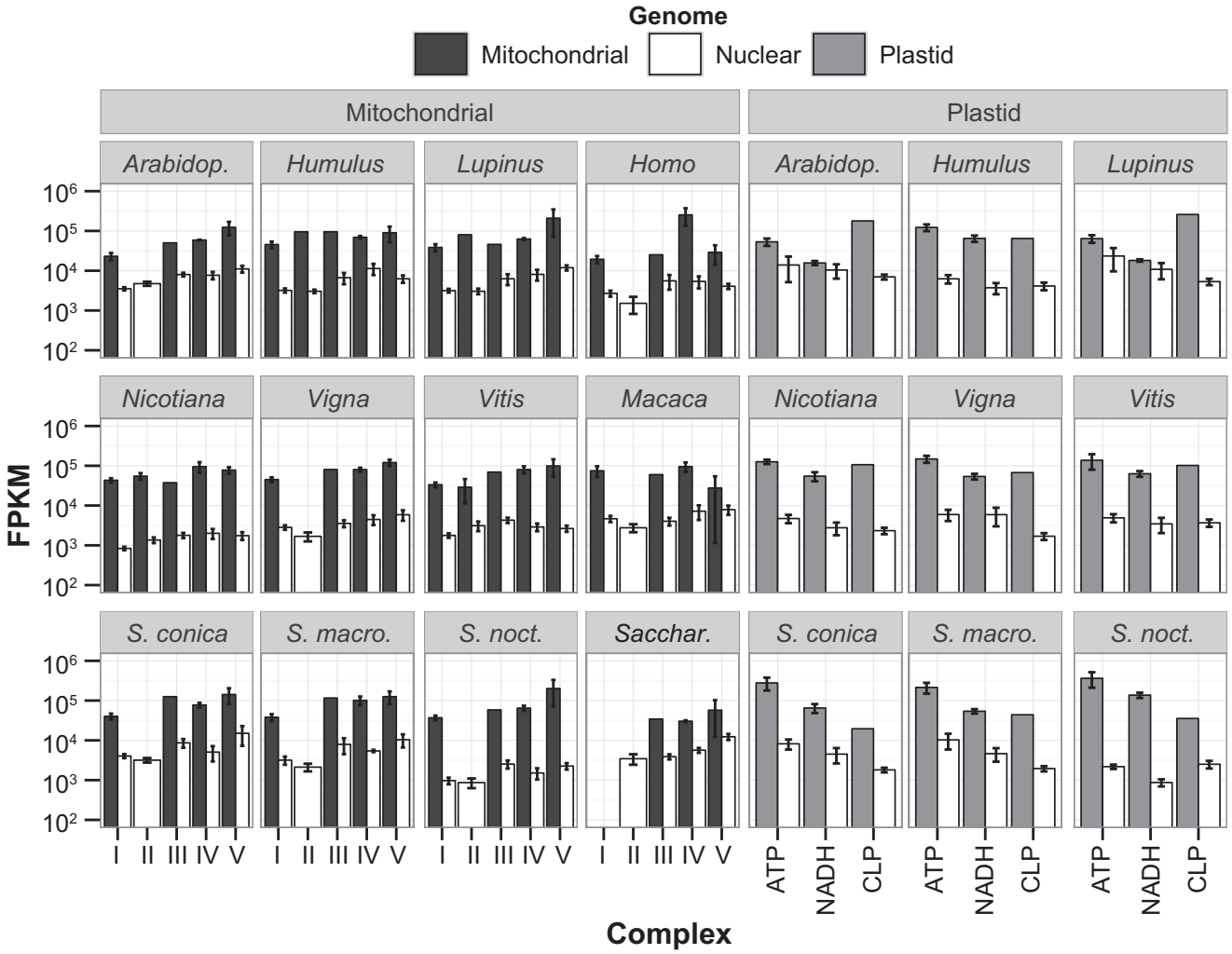

Fig. 3. Differences in transcript abundance between mt- and N-mt OXPHOS subunits among angiosperms, primates, and yeast. Note that no data are provided for complex I in yeast because the entire complex has been lost in that evolutionary lineage. For complex II (succinate dehydrogenase), *Nicotiana tabacum*, *Humulus lupulus*, *Lupinus albus*, and *Vitis vinifera* retain mt-encoded *sdh3* and *sdh4*, whereas *Arabidopsis thaliana*, *Silene conica*, *S. marcrodonta*, *S. noctiflora*, and *Vigna radiata* have transferred all complex II subunits to the nucleus (fig. 4). Differences in transcript abundance between plastid- and nuclear-encoded subunits of the plastid ATP synthase, NADH-plastoquinone oxidoreductase, and CLP protease complexes are also shown for nine angiosperms. *Homo*, *Macaca*, and *Sacchar.* are abbreviations of *Homo sapiens*, *Macaca mulatta*, and *Saccharomyces cerevisiae*. Estimates of transcript abundance are shown in supplementary data file S1, Supplementary Material online. Error bars show ± SEM.

should consistently outnumber nuclear ones. Complex V (ATP synthase) is the only case in which subunit stoichiometries deviate from 1:1. The subunits present at higher ratios within this complex do have higher transcript and protein abundances (supplementary fig. S6, Supplementary Material online), but the large excess of mt transcripts still exists after accounting for this imbalanced stoichiometry (supplementary fig. S7, Supplementary Material online).

One obvious contributing factor is that there are more mt genome copies per cell than nuclear genome copies. However, if genome copy number were solely driving this trend, then all mt-encoded genes should show similar, high levels of transcript abundance, which was not the case. There was a wide range of variation in transcript abundance within mt-encoded genes (e.g., a 36-fold range among mt genes in *Arabidopsis*; similar to that observed in animals; Nabholz et al. 2013). Moreover, the difference between plastid- and nuclear-encoded genes (figs. 3 and 5; supplementary fig. S8, Supplementary Material online) was similar in magnitude to the difference for mt and N-mt genes (18.5- versus 18.3-fold) even though plastid genomes are typically found at much higher copy numbers than mt genomes (Kumar et al. 2014; Raven 2015).

Despite differences in transcript abundance, relative mt and nuclear protein levels may be brought back into balance via post-transcriptional regulation. Support for this explanation comes from estimates of protein abundance in MitoCarta, an inventory of mammalian mt proteins (Pagliarini et al. 2008; Calvo et al. 2016), showing that mt OXPHOS proteins are not more abundant than nuclear ones (supplementary fig. S9, Supplementary Material online). Additionally, regulation of mt transcription is thought to be rather crude (Cantatore et al. 1987; Taanman 1999) and most regulatory mechanisms are likely post-transcriptional

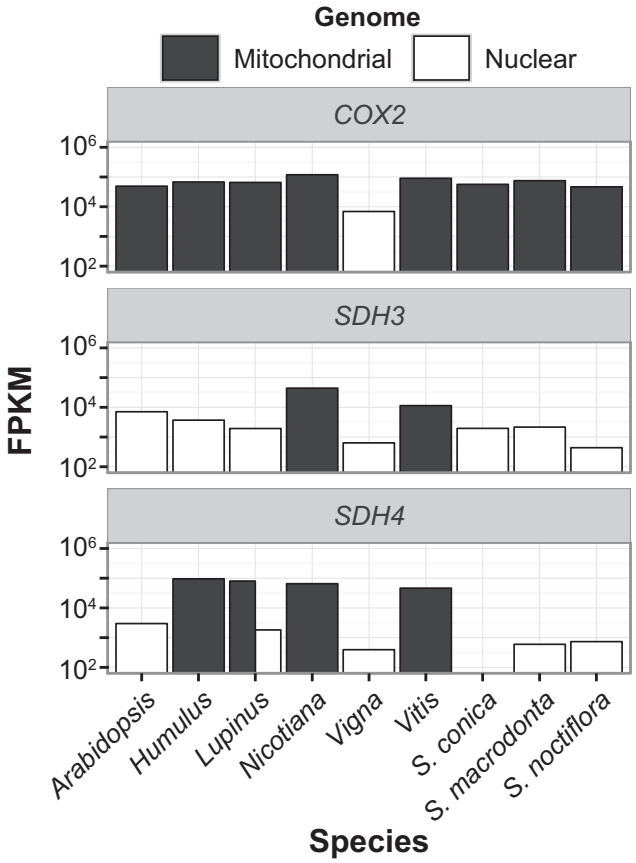

FIG. 4. Differences in transcript abundance between mt and N-mt OXPHOS subunits that have undergone intracellular gene transfer among angiosperms. Note that no reads mapped to the nuclear-encoded *sdh4* in *S. conica*. Both the mt- and nuclear-encoded copy of *sdh4* may be functional in *Lupinus* (see Materials and Methods).

(Woodson and Chory 2008). Mitochondrial transcript levels do change during development but are not necessarily correlated with protein levels (Steele et al. 1996; Giege et al. 2005; Woodson and Chory 2008). It has been shown that more mRNA molecules are present in rat mitochondria than can be physically translated (Garstka et al. 1994), which supports the notion that mt translation is less efficient than nuclear translation (Kim and Warner 1983; Woodson and Chory 2008). Inefficient mt translation might be due to increased ROS (reactive oxygen species) production and transcript damage in the mitochondria (Rep and Grivell 1996; Schwarze et al. 1998) or due to a history of reduced functional constraint and accumulation of weakly deleterious mutations in mt translational machinery (Brown et al. 1982; Lynch and Blanchard 1998; Barreto and Burton 2013; Sloan et al. 2014; Adrion et al. 2015; Pett and Lavrov 2015).

Excess mt transcripts may also play a functional role. For example, in some situations excess mt transcription may help to "prime" organelles for response to rapid changes in redox conditions or other environmental factors (Ulery et al. 1994; Matsuo and Obokata 2002; Woodson and Chory 2008). It was recently shown that mt OXPHOS subunits exhibit a much slower transcriptional response than N-mt OXPHOS subunits during mitochondrial biogenesis in yeast (Couvillion et al. 2016). In contrast, translational responses for both mt and N-mt OXPHOS subunits were rapid and synchronously regulated, such that abundances for both sets of proteins change in a coordinated fashion (Couvillion et al. 2016). Therefore, although mt genes do exhibit some levels of transcriptional response to environmental change (Flight et al. 2011), we propose that maintaining a large pool of excess mt transcripts is necessary to compensate for inefficient transcriptional regulation in the mitochondria so that mt translation can keep pace with that of the nucleus. These observations are also relevant to the hypothesis that organelles retain their own genomes in order to respond quickly to the unique redox state found in individual organelles and preserve a level of local control independent of the nucleus (Allen 2003, 2015). It is possible that having a glut of mRNA transcripts proximate to the inner mt membrane may enable post-transcriptional regulation of OXPHOS in response to local redox conditions.

Our findings that mt transcripts far outnumber nuclear-encoded transcripts have implications for the role of mitochondria in aging. Specifically, the balance of mt to nuclear-encoded OXPHOS proteins is significantly altered in both nematode and mammalian systems with extended lifespans (Houtkooper et al. 2013). Multiple lifespan-extension methods result in decreased concentrations of mt OXPHOS proteins relative to N-mt proteins, shifting the mt:nuclear protein balance from ~1:1 to up to ~1:75 depending on the longevity induction method used (including knocking down N-mt ribosomal protein expression itself; Houtkooper et al. 2013). If mt transcript abundance determines mt protein abundance (at least under some conditions), then reducing mt transcript levels may also cause increased lifespan. Therefore, examining OXPHOS transcript abundance in normal versus long-lived systems may provide insight into the contentious relationship between mitochondrial function and longevity (Wang and Hekimi 2015).

Materials and Methods

Transcript Abundance

To determine whether mt-encoded OXPHOS subunits are expressed more highly than nuclear-encoded OXPHOS subunits in other eukaryotes besides animals, publicly available RNA-seq data sets were downloaded from the Sequence Read Archive (SRA) at the National Center for Biotechnology Information (NCBI; supplementary table S1, Supplementary Material online). Our search was limited to libraries that were prepared using rRNA depletion strategies (i.e., Ribo-Zero) rather than poly-A selection because mt transcripts are often not polyadenylated and could be highly underrepresented in conventional RNA-seq libraries (Castandet et al. 2016). The overall data set used to investigate transcript abundance consisted of nine plant species, one yeast, and two primates (supplementary table S1, Supplementary Material online). For *Arabidopsis thaliana*, we examined three different data sets (supplementary table S1 and fig. S4, Supplementary Material online), which all showed the same general pattern,

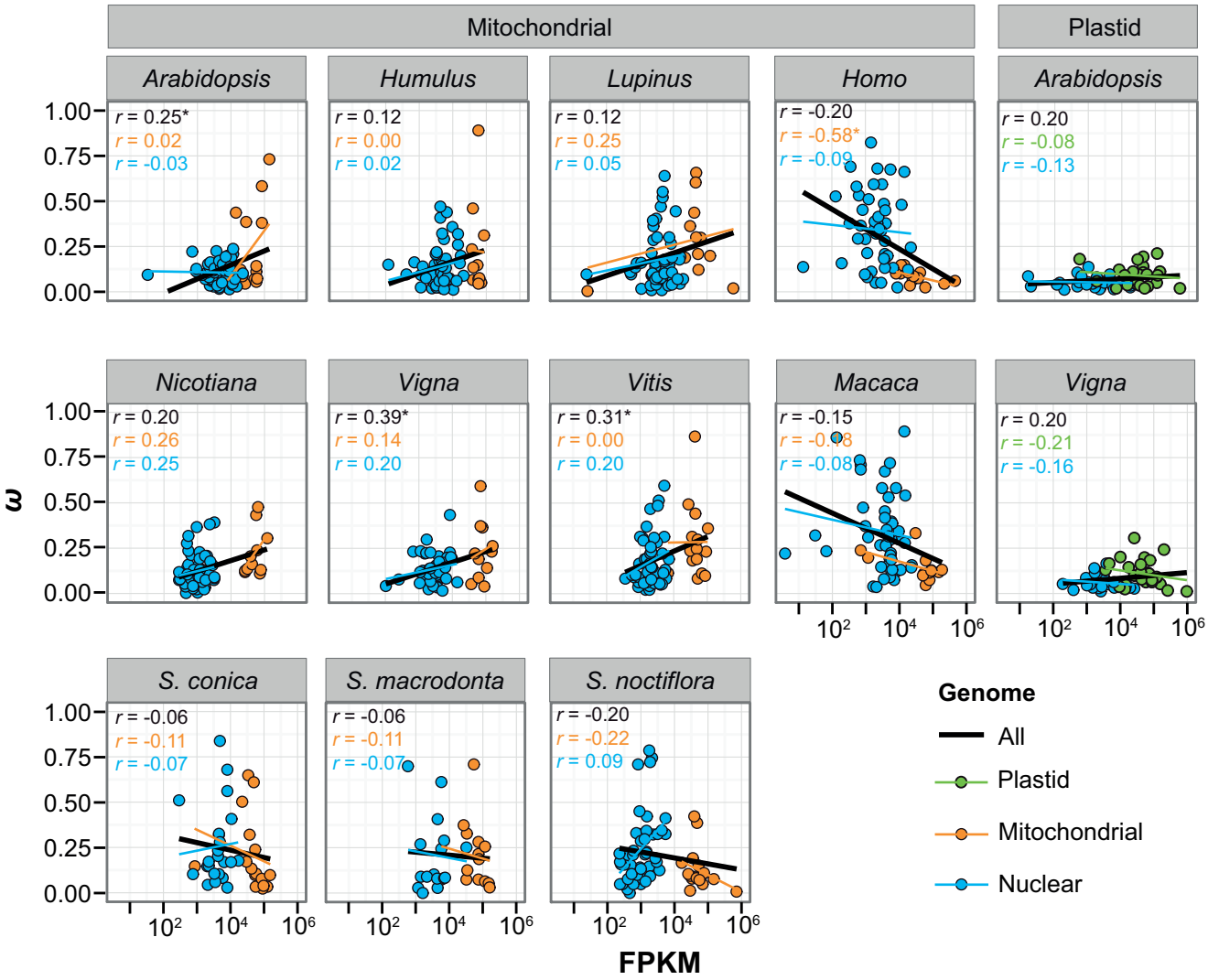

Fig. 5. Relationship between transcript abundance and ω for mt and N-mt OXPHOS subunits among angiosperms. Plastid- and nuclear-encoded subunits from the plastid ATP synthase, NADH-plastoquinone oxidoreductase, and photosystems I and II are also presented (see supplementary fig. S5, Supplementary Material online, for presentation of plastid data on an expanded *y* axis that more clearly depicts trends in this data set). Best-fit lines in black are for the total data set, including both genomes, whereas those in orange, blue, and green use solely mt-, nuclear-, or plastid-encoded encoded genes, respectively. The *y* axis was truncated at $\omega = 1$, excluding some outliers, although all data points were used when fitting lines and calculating statistics. All data points are available in supplementary data file S3, Supplementary Material online.

so the results from only one of these data sets are presented in the main text.

These eukaryotes possess different numbers of mt-encoded OXPHOS genes, which allowed us to investigate how transcript abundance changes based on whether a specific gene is encoded by the nuclear versus mt genome. We focused on three genes (*sdh3*, *sdh4*, and *cox2*) to specifically investigate the effects of intracellular gene transfer on expression. The *sdh3* and *sdh4* genes are encoded by the mt genome in *Vitis* and *Nicotiana*, and the nuclear genome in *Arabidopsis*, *Silene*, and *Vigna* (Alverson et al. 2011; Skippington et al. 2015). Mt genomes have not been sequenced from *Humulus* and *Lupinus*, so we investigated whether these genes are likely mt- or nuclear-encoded in these species. Using the presence of predicted N-terminal signal peptides targeted to the mitochondria, the results from a previous Southern-blot analysis (Adams et al. 2002), and molecular analyses of sequence divergence, we classified *SDH3* as nuclear-encoded in *Humulus* and *Lupinus*. While *sdh4* is likely mt-encoded in *Humulus*, we found a functional copy in both genomes of *Lupinus*, possibly representing a case of intracellular gene transfer "in action". For *cox2*, a single case of intracellular transfer to the nucleus is found in *Vigna*, whereas all other species have only a mt copy (Adams et al. 1999).

To calculate transcript abundance, transcriptomes were first assembled *de novo* for each RNA-seq data set using Trinity r20120608 with default settings. Nuclear- and mt-encoded OXPHOS genes were then identified from each set of contigs by using tBLASTx v2.2.29+ with OXPHOS subunit protein sequences from the *Arabidopsis* Mitochondrial Protein Database as queries. In some cases (*Vigna*, *Nicotiana*, *Arabidopsis*, and *Saccharomyces*), OXPHOS gene sequences were acquired from published mt genome sequences (Sugiyama et al. 2005; Alverson et al. 2011) or

publicly available databases (Sickmann et al. 2003; Heazlewood et al. 2004; The Arabidopsis Information Resource [TAIR], www.arabidopsis.org; last accessed 13 Aug 2015). Raw reads were then mapped to the OXPHOS sequences using Bowtie version 1.1.1 and transcript abundances estimated as FPKM values with RSEM v1.2.12 as implemented in Trinity version trinityrnaseq_r20140717 using the align_and_estimate_abundance.pl script (Langmead 2010; Li and Dewey 2011; Haas et al. 2013). Importantly, each read was only allowed to map to one OXPHOS sequence, and, in cases where reads mapped equally well to more than one sequence, a single sequence was chosen at random. Finally, for nuclear OXPHOS sequences with more than one isoform, FPKM values from each isoform were summed, resulting in a single FPKM value per OXPHOS subunit.

Expression of plastid genes and nuclear genes that encode plastid-targeted proteins were investigated in a similar manner using the same set of nine angiosperms. Briefly, *Arabidopsis* sequences were used as tBLASTx queries to identify genes in the other species that encoded subunits of the plastid ATP synthase, NADH-plastoquinone oxidoreductase, and CLP protease complexes. Estimates of gene expression were calculated as above for OXPHOS genes.

Signatures of Selection on OXPHOS Genes

Mt- and nuclear-encoded OXPHOS gene sequences were investigated for signatures of selection and relative mt substitution rate from across a diverse set of eukaryotes (table 1). In total, we analyzed data from 84 species in 15 different taxonomic groups from animals, plants, and fungi (supplementary table S2, Supplementary Material online). Sequences were acquired from either the KEGG database for a diverse range of eukaryotes (supplementary table S2, Supplementary Material online) or from the plant transcriptomes mentioned above that were used to estimate transcript abundance. For corals, publicly available transcriptomes and mt genomes were used to identify OXPHOS genes, and for four coral species mt sequences were acquired from congeners with available nuclear data. Coral OXPHOS genes were identified by using tBLASTx with OXPHOS protein sequences from either the published *Nematostella* or *Hydra* genomes as queries. For genes acquired from the KEGG database, sequences were obtained via the curated KEGG OXPHOS metabolic pathway (#00190) and its associated ortholog table for eukaryotes. For species with more than one entry per gene (e.g., multiple isoforms or splice variants of the same annotated gene), the first entry in the KEGG ortholog table was used. For angiosperms, RNA-edited sites in mtDNA-encoded genes were predicted using PREP-Mt with a cutoff value of 0.2 and edited sequences were used in all subsequent analyses (Mower 2005). Signal peptides in the N-mt genes were predicted using the TargetP 1.1 server for a subset of angiosperms with low and high mt substitution rate and were excluded to investigate their effects on the results (supplementary fig. S2, Supplementary Material online) (Emanuelsson et al. 2007).

To investigate a "control" set of nuclear-encoded genes that are not targeted to the mitochondria, genes in the KEGG glycolysis pathway (#00010) were used to generate parallel estimates of nuclear d_N and d_S. Glycolysis genes were chosen as a control for the OXPHOS analysis because they also play a role in energy metabolism, the lengths of the concatenated alignments were similar between OXPHOS and glycolysis data sets, and the rate of evolution was similar for the two data sets. Glycolysis genes were extracted from the KEGG database following the same methods as for OXPHOS genes.

Species were classified taxonomically into 15 different groups (supplementary table S2, Supplementary Material online) and codon-based alignments were generated for each homologous OXPHOS gene for each taxonomic group using MUSCLE v3.7 (Edgar 2004) as implemented in MEGA v6.06 (Tamura et al. 2013) and via manually checking the alignments for alignment errors. Nucleotide sequences from the resulting alignments were then concatenated for each taxonomic group, with mt- and nuclear-encoded genes concatenated separately. Codeml in PAML v4.8 (Yang 2007) was used with an F1 × 4 model to estimate separate d_N and d_S values for concatenated mt and nuclear genes from each of the taxonomic groups. Although a single d_N, d_S, and ω value was estimated for each alignment (model = 0 in codeml), estimates were also generated for each species individually (model = 1 in codeml) in order to exclude species with a saturated number of substitutions in either genome (i.e., $d_S \geq 3$), and provide another measure of the correlation between relative mt selection and substitution rate (supplementary fig. S1, Supplementary Material online; see below). For each taxonomic group, an input phylogenetic tree was used based on currently accepted patterns of relatedness for the species in question (supplementary table S3, Supplementary Material online). For data sets from amphibians and fungi (*Aspergillus*), only two species were used, so the pairwise model of divergence (runmode = −2 in codeml) was employed in PAML. Statistical correlations between $d_{S\text{-mt}}/d_{S\text{-nuc}}$ and ω_{mt}/ω_{nuc} were investigated via regression analyses on log-transformed values using the lm function as implemented in R v2.15.0 (R Development Core Team 2012). We also explicitly accounted for phylogenetic history in this correlation by employing phylogenetic independent contrasts using the ape package in R (Paradis et al. 2004).

Effects of Expression on ω

To determine if a negative relationship existed between transcript abundance and ω in angiosperms, as it does in animals (Nabholz et al. 2013), ω values were calculated for individual OXPHOS subunits using PAML for each of the nine angiosperms where FPKM was also calculated (model = 1). Genes with saturated (> 3) or small (< 0.01) d_S or d_N values were excluded, and regression analyses between log-transformed FPKM values and ω values were calculated using the lm function in R. This analysis was also performed for two primate species, *Homo sapiens* and *Macaca mulatta* (using *Callithrix jacchus* sequences as an outgroup in PAML analyses), to confirm the previous negative relationship observed in animals. Finally, to determine if transcript abundance correlates with ω in plastid genes, we investigated subunits of the plastid ATP synthase, NADH-plastoquinone oxidoreductase, and

photosystems I and II in *Vitis*, *Arabidopsis*, and *Vigna*. Photosystem sequences were acquired by using those in the KEGG photosynthesis metabolic pathway (#00195) and expression was quantified as above. ω values were estimated as above, with *Vitis* as an outgroup, and therefore correlations were only performed for *Vigna* and *Arabidopsis*.

Estimating OXPHOS Protein Abundances

The Human MitoCarta 2.0 database was downloaded from http://www.broadinstitute.org/scientific-community/science/programs/metabolic-disease-program/publications/mitocarta/mitocarta-in-0; (last accessed 2 Sept 2016) and mt and nuclear OXPHOS subunits were extracted from the "Tmito" genes predicted to be localized to mitochondria (Calvo et al. 2016). Protein abundance was then estimated using "MSMS Total Intensity", which is based on pooling spectra from 14 different mouse tissues.

Data Access

Estimates of transcript abundance are available as FPKM values for each species examined, for both OXPHOS and plastid genes, in supplementary data file S1, Supplementary Material online. Concatenated alignments that were used to estimate d_N, d_S, and ω of OXPHOS (mt and nuclear), glycolysis, and plastid genes for each taxonomic group are available via FigShare (https://figshare.com/articles/SupplData_sequences_used_for_dNdS_Havird_Sloan_MBE/3591033), last accessed on 2 Sept 2016 and d_N, d_S, and ω values are available in supplementary data file S2, Supplementary Material online. Estimates of transcript abundance and d_N, d_S, and ω for *Homo*, *Macaca*, and angiosperm data sets are available in supplementary data file S3, Supplementary Material online. Sets of contigs that encompass the assembled transcriptomes are available as described for *Silene* (supplementary table S1, Supplementary Material online), and will be made available upon request for other taxa (along with any other data).

Supplementary Material

Supplementary figures S1–S9, tables S1–S3 and supplementary data files S1–S3 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Author Contributions

J.C.H. and D.B.S. designed the research and wrote the article and J.C.H. performed the research and analyzed the data.

Acknowledgments

We thank Rachel Muller and three anonymous reviewers for valuable comments on an earlier version of this article and members of the labs of DBS, Karyn Hamilton, and Benjamin Miller for helpful discussion. This work was supported by a Division of Molecular and Cellular Biosciences grant at the National Science Foundation grant (MCB 1412260) and a National Institutes of Health Ruth L. Kirschstein National Research Service Award (F32GM116361).

References

Adams KL, Palmer JD. 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol Phylogenet Evol.* 29:380–395.

Adams KL, Qiu YL, Stoutemyer M, Palmer JD. 2002. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proc Natl Acad Sci U S A.* 99:9905–9912.

Adams KL, Song K, Roessler PG, Nugent JM, Doyle JL, Doyle JJ, Palmer JD. 1999. Intracellular gene transfer in action: dual transcription and multiple silencings of nuclear and mitochondrial cox2 genes in legumes. *Proc Natl Acad Sci U S A.* 96:13863–13868.

Adrion JR, White PS, Montooth KL. 2015. The roles of compensatory evolution and constraint in aminoacyl tRNA synthetase evolution. *Mol Biol Evol.* 33:152–161.

Allen JF. 2003. Why chloroplasts and mitochondria contain genomes. *Comp Funct Genomics* 4:31–36.

Allen JF. 2015. Why chloroplasts and mitochondria retain their own genomes and genetic systems: colocation for redox regulation of gene expression. *Proc Natl Acad Sci U S A.* 112:10231–10238.

Alverson AJ, Zhuo S, Rice DW, Sloan DB, Palmer JD. 2011. The mitochondrial genome of the legume Vigna radiata and the analysis of recombination across short mitochondrial repeats. *Plos One* 6:e16404.

Barreto FS, Burton RS. 2013. Evidence for compensatory evolution of ribosomal proteins in response to rapid divergence of mitochondrial rRNA. *Mol Biol Evol.* 30:310–314.

Beck EA, Thompson AC, Sharbrough J, Brud E, Llopart A. 2015. Gene flow between *Drosophila yakuba* and *Drosophila santomea* in subunit V of cytochrome c oxidase: a potential case of cytonuclear cointrogression. *Evolution* 69:1973–1986.

Boore JL. 1999. Animal mitochondrial genomes. *Nucleic Acids Res.* 27:1767–1780.

Brown WM, George M, Wilson AC. 1979. Rapid evolution of animal mitochondrial-DNA. *Proc Natl Acad Sci U S A.* 76:1967–1971.

Brown WM, Prager EM, Wang A, Wilson AC. 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution. *J Mol Evol.* 18:225–239.

Burger G, Gray MW, Forget L, Lang BF. 2013. Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. *Genome Biol Evol.* 5:418–438.

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(Suppl 6):S14–S24.

Burton RS, Pereira RJ, Barreto FS. 2013. Cytonuclear genomic interactions and hybrid breakdown. *Annu Rev Ecol Evol Syst.* 44:281–302.

Calvo SE, Clauser KR, Mootha VK. 2016. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res.* 44:D1251–D1257.

Cantatore P, Flagella Z, Fracasso F, Lezza AM, Gadaleta MN, de Montalvo A. 1987. Synthesis and turnover rates of four rat liver mitochondrial RNA species. *FEBS Lett.* 213:144–148.

Castandet B, Hotto AM, Strickler SR, Stern DB. 2016. ChloroSeq, an optimized chloroplast RNA-Seq bioinformatic pipeline, reveals remodeling of the organellar transcriptome under heat stress. *G3* Advance Access published July 8, 2016, doi: 10.1534/g3.116.030783.

Chamary JV, Parmley JL, Hurst LD. 2006. Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nat Rev Genet.* 7:98–108.

Christensen AC. 2014. Genes and junk in plant mitochondria-repair mechanisms and selection. *Genome Biol Evol.* 6:1448–1453.

Cooper BS, Burrus CR, Ji C, Hahn MW, Montooth KL. 2015. Similar efficacies of selection shape mitochondrial and nuclear genes in both *Drosophila melanogaster* and *Homo sapiens*. *G3 (Bethesda)* 5:2165–2176.

Couvillion MT, Soto IC, Shipkovenska G, Churchman LS. 2016. Synchronized mitochondrial and cytosolic translation programs. *Nature* 533:499–503.

Denver DR, Morris K, Lynch M, Vassilieva LL, Thomas WK. 2000. High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. *Science* 289:2342–2344.
Dowling DK. 2014. Evolutionary perspectives on the links between mitochondrial genotype and disease phenotype. *Biochim Biophys Acta* 1840:1393–1403.
Dowling DK, Friberg U, Hailer F, Arnqvist G. 2007. Intergenomic epistasis for fitness: within-population interactions between cytoplasmic and nuclear genes in *Drosophila melanogaster*. *Genetics* 175:235–244.
Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.
Emanuelsson O, Brunak S, von Heijne G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc.* 2:953–971.
Flight PA, Nacci D, Champlin D, Whitehead A, Rand DM. 2011. The effects of mitochondrial genotype on hypoxic survival and gene expression in a hybrid population of the killifish, *Fundulus heteroclitus*. *Mol Ecol.* 20:4503–4520.
Garstka HL, Facke M, Escribano JR, Wiesner RJ. 1994. Stoichiometry of mitochondrial transcripts and regulation of gene expression by mitochondrial transcription factor A. *Biochem Biophys Res Commun.* 200:619–626.
Giege P, Sweetlove LJ, Cognat V, Leaver CJ. 2005. Coordination of nuclear and mitochondrial genome expression during mitochondrial biogenesis in Arabidopsis. *Plant Cell* 17:1497–1512.
Gong L, Olson M, Wendel JF. 2014. Cytonuclear evolution of rubisco in four allopolyploid lineages. *Mol Biol Evol.* 31:2624–2636.
Gray MW, Burger G, Lang BF. 1999. Mitochondrial evolution. *Science* 283:1476–1481.
Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc.* 8:1494–1512.
Havird JC, Whitehill NS, Snow CD, Sloan DB. 2015. Conservative and compensatory evolution in oxidative phosphorylation complexes of angiosperms with highly divergent rates of mitochondrial genome evolution. *Evolution* 69:3069–3081.
Heazlewood JL, Tonti-Filippini JS, Gout AM, Day DA, Whelan J, Millar AH. 2004. Experimental analysis of the *Arabidopsis* mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins. *Plant Cell* 16:241–256.
Hill GE. 2015. Mitonuclear ecology. *Mol Biol Evol.* 32:1917–1927.
Hjort K, Goldberg AV, Tsaousis AD, Hirt RP, Embley TM. 2010. Diversity and reductive evolution of mitochondria among microbial eukaryotes. *Philos Trans R Soc Lond B Biol Sci.* 365:713–727.
Houtkooper RH, Mouchiroud L, Ryu D, Moullan N, Katsyuba E, Knott G, Williams RW, Auwerx J. 2013. Mitonuclear protein imbalance as a conserved longevity mechanism. *Nature* 497:451–457.
Karnkowska A, Vacek V, Zubacova Z, Treitli SC, Petrzelkova R, Eme L, Novak L, Zarsky V, Barlow LD, Herman EK, et al. 2016. A eukaryote without a mitochondrial organelle. *Curr Biol.* 26:1274–1284.
Kim CH, Warner JR. 1983. Messenger RNA for ribosomal proteins in yeast. *J Mol Biol.* 165:79–89.
Kumar RA, Oldenburg DJ, Bendich AJ. 2014. Changes in DNA damage, molecular integrity, and copy number for plastid DNA and mitochondrial DNA during maize development. *J Exp Bot.* 65:6425–6439.
Langmead B. 2010. Aligning short sequencing reads with Bowtie. *Curr Protoc Bioinformatics* Chapter 11:Unit 11.7.
Lavrov DV. 2007. Key transitions in animal evolution: a mitochondrial DNA perspective. *Integr Comp Biol.* 47:734–743.
Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323.
Li J, Zhang Z, Vang SR, Yu J, Wong GKS, Wang J. 2009. Correlation between Ka/Ks and Ks is related to substitution model and evolutionary lineage. *J Mol Evol.* 68:414–423.
Lynch M, Blanchard JL. 1998. Deleterious mutation accumulation in organelle genomes. *Genetica* 102–103:29–39.
Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, Dickinson WJ, Okamoto K, Kulkarni S, Hartl DL, et al. 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci U S A.* 105:9272–9277.
Maier UG, Zauner S, Woehle C, Bolte K, Hempel F, Allen JF, Martin WF. 2013. Massively convergent evolution for ribosomal protein gene content in plastid and mitochondrial genomes. *Genome Biol Evol.* 5:2318–2329.
Matsuo M, Obokata J. 2002. Dual roles of photosynthetic electron transport in photosystem I biogenesis: light induction of mRNAs and chromatic regulation at post-mRNA level. *Plant Cell Physiol.* 43:1189–1197.
Meiklejohn CD, Holmbeck MA, Siddiq MA, Abt DN, Rand DM, Montooth KL. 2013. An Incompatibility between a mitochondrial tRNA and its nuclear-encoded tRNA synthetase compromises development and fitness in *Drosophila*. *PLoS Genet.* 9:e1003238.
Mower JP. 2005. PREP-Mt: predictive RNA editor for plant mitochondrial genes. *BMC Bioinformatics* 6:96.
Mower JP, Bonen L. 2009. Ribosomal protein L10 is encoded in the mitochondrial genome of many land plants and green algae. *BMC Evol Biol.* 9:265.
Mower JP, Touzet P, Gummow JS, Delph LF, Palmer JD. 2007. Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. *BMC Evol Biol.* 7:135.
Nabholz B, Ellegren H, Wolf JB. 2013. High levels of gene expression explain the strong evolutionary constraint of mitochondrial protein-coding genes. *Mol Biol Evol.* 30:272–284.
Neiman M, Taylor DR. 2009. The causes of mutation accumulation in mitochondrial genomes. *Proc Biol Sci.* 276:1201–1209.
Nielsen R. 2005. Molecular signatures of natural selection. *Annu Rev Genet.* 39:197–218.
Osada N, Akashi H. 2012. Mitochondrial-nuclear interactions and accelerated compensatory evolution: evidence from the primate cytochrome C oxidase complex. *Mol Biol Evol.* 29:337–346.
Pagliarini DJ, Calvo SE, Chang B, Sheth SA, Vafai SB, Ong SE, Walford GA, Sugiana C, Boneh A, Chen WK, et al. 2008. A mitochondrial protein compendium elucidates complex I disease biology. *Cell* 134:112–123.
Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.
Petersen J, Ludewig AK, Michael V, Bunk B, Jarek M, Baurain D, Brinkmann H. 2014. *Chromera velia*, endosymbioses and the rhodoplex hypothesis–plastid evolution in cryptophytes, alveolates, stramenopiles, and haptophytes (CASH lineages). *Genome Biol Evol.* 6:666–684.
Pett W, Lavrov D. 2015. Mito-nuclear interactions in the evolution of animal mitochondrial tRNA metabolism. *Genome Biol Evol.* 7:2089–2101.
Popadin KY, Nikolaev SI, Junier T, Baranova M, Antonarakis SE. 2013. Purifying selection in mammalian mitochondrial protein-coding genes is highly effective and congruent with evolution of nuclear genes. *Mol Biol Evol.* 30:347–355.
R Development Core Team. 2012. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
Raven JA. 2015. Implications of mutation of organelle genomes for organelle function and evolution. *J Exp Bot.* 66:5639–5650.
Rep M, Grivell LA. 1996. The role of protein degradation in mitochondrial function and biogenesis. *Curr Genet.* 30:367–380.
Schwarze SR, Weindruch R, Aiken JM. 1998. Oxidative stress and aging reduce COX I RNA and cytochrome oxidase activity in *Drosophila*. *Free Radic Biol Med.* 25:740–747.
Sickmann A, Reinders J, Wagner Y, Joppich C, Zahedi R, Meyer HE, Schonfisch B, Perschil I, Chacinska A, Guiard B, et al. 2003. The proteome of *Saccharomyces cerevisiae* mitochondria. *Proc Natl Acad Sci U S A.* 100:13207–13212.
Skippington E, Barkman TJ, Rice DW, Palmer JD. 2015. Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all nad genes. *Proc Natl Acad Sci U S A.* 112:E3515–E3524.

Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR. 2012. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLoS Biol.* 10:e1001241.

Sloan DB, Oxelman B, Rautenberg A, Taylor DR. 2009. Phylogenetic analysis of mitochondrial substitution rate variation in the angiosperm tribe Sileneae. *BMC Evol Biol.* 9:260.

Sloan DB, Triant DA, Wu M, Taylor DR. 2014. Cytonuclear interactions and relaxed selection accelerate sequence evolution in organelle ribosomes. *Mol Biol Evol.* 31:673–682.

Smith DR, Keeling PJ. 2015. Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. *Proc Natl Acad Sci U S A.* 112:10177–10184.

Steele DF, Butler CA, Fox TD. 1996. Expression of a recoded nuclear gene inserted into yeast mitochondrial DNA is limited by mRNA-specific translational activation. *Proc Natl Acad Sci U S A.* 93:5253–5257.

Stoletzki N, Eyre-Walker A. 2011. The positive correlation between dN/dS and dS in mammals Is due to runs of adjacent substitutions. *Mol Biol Evol.* 28:1371–1380.

Sugiyama Y, Watase Y, Nagase M, Makita N, Yagura S, Hirai A, Sugiura M. 2005. The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. *Mol Genet Genomics* 272:603–615.

Taanman JW. 1999. The mitochondrial genome: structure, transcription, translation and replication. *Biochim Biophys Acta* 1410:103–123.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 30:2725–2729.

Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet.* 5:123–135.

Ulery TL, Jang SH, Jaehning JA. 1994. Glucose repression of yeast mitochondrial transcription: kinetics of derepression and role of nuclear genes. *Mol Cell Biol.* 14:1160–1170.

Vallender EJ, Lahn BT. 2007. Uncovering the mutation-fixation correlation in short lineages. *BMC Evol Biol.* 7:168.

van der Sluis EO, Bauerschmitt H, Becker T, Mielke T, Frauenfeld J, Berninghausen O, Neupert W, Herrmann JM, Beckmann R. 2015. Parallel structural evolution of mitochondrial ribosomes and OXPHOS complexes. *Genome Biol Evol.* 7:1235–1251.

Wang Y, Hekimi S. 2015. Mitochondrial dysfunction and longevity in animals: untangling the knot. *Science* 350:1204–1207.

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

Wolfe KH, Li WH, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci U S A.* 84:9054–9058.

Woodson JD, Chory J. 2008. Coordination of gene expression between organellar and nuclear genomes. *Nat Rev Genet.* 9:383–395.

Wyckoff GJ, Malcom CM, Vallender EJ, Lahn BT. 2005. A highly unexpected strong correlation between fixation probability of nonsynonymous mutations and mutation rate. *Trends Genet.* 21:381–385.

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

Zhang F, Broughton RE. 2013. Mitochondrial-nuclear interactions: compensatory evolution or variable functional constraint among vertebrate oxidative phosphorylation genes?. *Genome Biol Evol.* 5:1781–1791.

Zhang J, Yang JR. 2015. Determinants of the rate of protein sequence evolution. *Nat Rev Genet.* 16:409–420.