Molecular Phylogenetics and Evolution 49 (2008) 827–831

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

# Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants

Guy Drouin *, Hanane Daoud, Junnan Xia

*Département de biologie et Centre de recherche avancée en génomique environnementale, Université d'Ottawa, 30 Marie Curie, Ottawa, Ont., Canada K1N 6N5*

ARTICLE INFO

*Article history:*
Received 17 March 2008
Revised 19 August 2008
Accepted 13 September 2008
Available online 21 September 2008

*Keywords:*
Synonymous
Substitutions
Angiosperms
Gnetales
Gymnosperms

ABSTRACT

Previous studies have estimated that, in angiosperms, the synonymous substitution rate of chloroplast genes is three times higher than that of mitochondrial genes and that of nuclear genes is twelve times higher than that of mitochondrial genes. Here we used 12 genes in 27 seed plant species to investigate whether these relative rates of substitutions are common to diverse seed plant groups. We find that the overall relative rate of synonymous substitutions of mitochondrial, chloroplast and nuclear genes of all seed plants is 1:3:10, that these ratios are 1:2:4 in gymnosperms but 1:3:16 in angiosperms and that they go up to 1:3:20 in basal angiosperms. Our results show that the mitochondrial, chloroplast and nuclear genomes of seed plant groups have different synonymous substitutions rates, that these rates are different in different seed plant groups and that gymnosperms have smaller ratios than angiosperms.


## 1. Introduction

Ever since the concept of a molecular clock was proposed, numerous studies have addressed its existence, speed and universality (Zuckerkandl and Pauling, 1965; Easteal et al., 1995; Graur and Li, 2000). One of the interesting findings of earlier studies was that the synonymous substitution rate of mammalian mitochondrial genes was about 10 times higher than that of mammalian nuclear genes (Brown et al., 1979, 1982; Miyata et al., 1982; Easteal et al., 1995; Graur and Li, 2000). This observation led to the suggestion that the higher substitution rate observed in mitochondrial genes was the result of a higher mutation rate due to presence of oxygen radicals in mitochondria. However, the subsequent observation that the mitochondrial genes of angiosperms (i.e., monocots and eudicots) evolve about 12 times more slowly than their nuclear genes cast doubts onto this hypothesis (Wolfe et al., 1987, 1989; Graur and Li, 2000).

Here, we revisit the studies of Wolfe and colleagues (Wolfe et al., 1987, 1989) where they observed that the synonymous rate of evolution of the mitochondrial, chloroplast and nuclear genes of angiosperms had ratios of 1:3:12. That is, the synonymous substitution rate of chloroplast genes of angiosperms species is three times higher than that of their mitochondrial genes and that of their nuclear genes nuclear genes are 12 times higher than that of their mitochondrial genes. Given the paucity of plant sequences available at the time, their results were based on a limited number of diverse genes from diverse species comparisons. Although synonymous rates are often similar in different genes, it is preferable to use the same genes in different species to eliminate gene to gene variation (Graur and Li, 2000). Similarly, making different species comparisons also introduces more variation in the rates inferred because the different divergence times used to calculate these rates are still uncertain (Wolfe et al., 1987, 1989). Furthermore, their data set was composed uniquely of monocot and eudicot sequences. It is therefore of interest to determine whether their results extend to other seed plant groups.

In this study, we compare the same 12 genes in the same 27 species. This not only insures that the same genes are compared in different species but also allows us to compare relative evolutionary rates without having to know the divergence times of the species being compared. The later is a significant advantage because estimating the divergence times of seed plant species is complex and still the subject of much debate (Soltis et al., 2002; Magallon and Sanderson, 2005). A disadvantage of this approach is that one has to have single copy genes from all three plant genomes (mitochondrial, chloroplast and nuclear) from the same (or closely related) species. Fortunately, this is no longer a problem with mitochondrial and chloroplast genes because the mitochondrial and chloroplast genes and genomes of many plant species have now been sequenced (mainly to study plant phylogenetic relationships). Furthermore, two sets of single copy (orthologous) plant nuclear genes are available from a relatively large sampling of seed plant species: the *rpb1* and *rpb2* genes data sets coding for the largest and second largest subunit of RNA polymerase II, respectively (Nickerson and Drouin, 2004; Oxelman et al., 2004;

* Corresponding author. Fax: +1 613 562 5486.
*E-mail address:* gdrouin@science.uottawa.ca (G. Drouin).

1055-7903/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2008.09.009

Hajibabaei et al., 2006; Luo et al., 2007). Two sets of single copy (orthologous) plant nuclear genes are also available from a number of angiosperm plant species: the *phyA* and *phyC* genes data sets coding for phytochrome A and C, respectively, (Mathews and Donoghue, 1999). Here, we used these existing sequences, as well as 11 new *rpb1* sequences, to investigate the relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants.

## 2. Materials and methods

### 2.1. Species, sequences and groups

Since the limiting factor for our analyses was the availability of nuclear (orthologous) single copy genes, we selected the 27 species for which we had *rpb1* genes sequences. This gene is particularly useful for our purposes because these sequences each contain over 3000 nucleotides of coding sequence. These 27 *rpb1* sequences are composed of previously published 16 cDNA sequences and 11 newly sequenced *rpb1* cDNA sequences (Nickerson and Drouin, 2004; Hajibabaei et al., 2006). The 11 new cDNA *rpb1* sequences are those of: *Asparagus officinialis*, *Beta vulgaris*, *Ceratophyllum demersum*, *Drimys winteri*, *Illicium parviflorum*, *Liriodendron tulipifera*, *Nicotiana tabacum*, *Papaver orientalis*, *Persea americanum*, *Pisum sativum* and *Saruma henryi*. They were cloned and sequenced as described in Nickerson and Drouin (2004) and have been deposited in GenBank under Accession Nos. EU543182–EU543192, respectively.

Sequences of the other 11 genes for all 27 species (or species closely related to them) were obtained from GenBank (http://www.ncbi.nlm.nih.gov/; Table 1, Supplemental Tables 1–3) except for the maize and rice *rpb2* genes which were obtained at http://tigrblast.tigr.org/ using the *Dioscorea sansibarensis rpb2* sequence (Accession No. AY563268) as a query. The other genes we used where chosen because their sequences are available for the plant species for which we had *rpb1* sequences. In asterid species having duplicated *rpb2* genes, we used the d clade sequences (as defined by Oxelman et al., 2004). Sequences were edited and aligned using BioEdit (Hall, 1999). This sequence editor uses ClustalW to align sequences and the alignments were performed using amino acid sequences (Thompson et al., 1994).

Given recent results in seed plant phylogenies, we divided the 27 species into seven groups. The first two groups were composed of 17 angiosperm species and 10 gymnosperm species. The angiosperm group was further divided into eudicots, monocots and basal angiosperms because this reflects their currently accepted phylogeny, with basal angiosperms being the sister group to eudicots and monocots (Table 2; Kuzoff and Gasser, 2000; Jansen et al., 2007; Moore et al., 2007). The gymnosperm group was further divided into gnetales and other gymnosperms because gnetales are a monophyletic group with genes that evolve faster than those of other gymnosperms (Table 2; Kuzoff and Gasser, 2000; Burleigh and Mathews, 2004; Hajibabaei et al., 2006).

### 2.2. Synonymous Substitutions

Numbers of synonymous substitutions per synonymous sites ($K_s$) and nonsynonymous substitutions per nonsynonymous sites ($K_a$) were calculated using programs from the PAML package version 4b (Yang, 2007). The YN00 program was used to calculate $K_s$ and $K_a$ using the methods of Li et al. (1985) whereas the CODEML program was used to calculate $K_s$ and $K_a$ using the maximum likelihood (ML) method (using the options seqtype = 1, runmode = −2 and CodonFreq = 2 in the codeml.ctl files). In all analyses, we calculated within group averages of all $(n \times (n-1))/2$ pairwise sequence comparisons within each group, as well as the variance and standard errors, using Excel 2003 (Microsoft Corporation, Redmond, Washington). Thus, the calculated values are based on as little as a single pairwise comparison (in the cases where we were missing one of the three gnetales species) to as many as 351 pairwise comparisons (in the cases where we calculated values for all 27 seed plant species). This affected the standard errors calculated, with smaller data sets having relatively larger standard errors. Note that we assumed that the effect of RNA editing on $K_s$ and $K_a$ estimates of mitochondrial genes are negligible. *P*-values of between group differences were calculated using one-tail *z*-tests as implemented in Excel 2003.

**Table 1**
List of genes, alignment lengths and gene lengths

| |
|---|
| Mitochondrial (6519) |
| *atpA*, ATPase alpha subunit (∼1250) |
| *cox1*, cytochrome oxidase subunit 1 (∼1400) |
| *matR*, maturase R (∼1850) |
| Chloroplastic (8562) |
| *atpB*, ATP synthase beta subunit (∼1450) |
| *matK*, maturase K (∼1500) |
| *psaA*, photosystem I subunit A (∼2150) |
| *psbB*, photosystem II CP47 protein (∼1500) |
| *rbcL*, ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (∼1400) |
| Nuclear (13551) |
| *phyA*, phytochrome A (∼1200) |
| *phyC*, phytochrome C (∼1200) |
| *rpb1*, RNA polymerase II largest subunit (∼3050) |
| *rpb2*, RNA polymerase II second largest subunit (∼1650) |

Notes. The number after the name of each genome is the nucleotide length of the sequence alignment used for this genome. The number after the name of each gene is the approximate number of nucleotides this gene has in most of the species we analyzed.

**Table 2**
List of species and groups

| |
|---|
| *Angiosperms* |
| Eudicots |
| *Arabidopsis thaliana* |
| *Beta vulgaris* |
| *Nicotiana tabacum* |
| *Papaver orientalis* |
| *Pisum sativum* |
| Monocots |
| *Asparagus officinalis* |
| *Oryza sativa* |
| *Zea mays* |
| Basal |
| *Amborella trichopoda* |
| *Ceratophyllum demersum* |
| *Drimys winteri* |
| *Illicium parviflorum* |
| *Liriodendron tulipifera* |
| *Magnolia soulangeana* |
| *Nymphaea odorata* |
| *Persea americanum* |
| *Saruma henryi* |
| *Gymnosperms* |
| Gnetales |
| *Ephedra viridis* |
| *Gnetum gnemon* |
| *Welwitschia mirabilis* |
| Other gymnosperms |
| *Cycas revoluta* |
| *Ginkgo biloba* |
| *Pinus nigra* |
| *Podocarpus macrophyllus* |
| *Taxus canadensis* |
| *Thuja occidentalis* |
| *Zamia muricata* |

## 3. Results

### 3.1. Sequences and alignments

Supplemental Tables 1–3 show that the data set we used is almost complete. Out of a total of 324 genes (12 genes × 27 species) the only sequences that are not available are those of three mitochondrial *matR* genes (for *Ephedra viridis*, *Taxus canadensis* and *Thuja occidentalis*), the nuclear *rpb1* gene of *Gnetum gnemon*, the nuclear *rpb2* gene of *C. demersum*, the *phyC* gene of *N. tabacum* and the *phyA* and *phyC* genes of gymnosperms (because these genes are specific to angiosperms).

Table 1 shows that although most sequences are partial sequences (i.e., they do not contain the full length of the coding region) most $K_s$ calculations were performed on coding regions containing about 4500 mitochondrial nucleotides, 8000 chloroplast nucleotides and 7100 nuclear nucleotides.

### 3.2. Synonymous and nonsynonymous substitutions

We used both the methods of Li et al. (1985) and the ML method implemented in PAML (Yang, 2007) to calculate $K_s$ and $K_a$. The first method was used so that our results could be compared to previous studies whereas the second method was used because it is considered to be the most accurate method currently available (Yang and Nielsen, 2000; Yang, 2007). In particular, the ML method is much better at dealing with multiple substitutions between distantly related sequences (Muse, 1996; Yang and Nielsen, 2000). This feature is essential to our study because we compare sequences as distantly related as those of gymnosperms and angiosperms. Below, unless otherwise noted, we only discuss results obtained with the ML method.

Supplemental Table 4a–c show that $K_s$ values for different genes within the same genome are very similar and do not vary by more than three-fold. For example, the average $K_s$ values for all angiosperms vary from 0.348 to 0.605 for chloroplast genes, from 0.066 to 0.184 for mitochondrial genes and from 2.043 to 4.162 for nuclear genes (Supplemental Table 4a–c). This is similar to what has been observed in the nuclear protein coding genes of mammals where synonymous rates have been observed to vary by at most three-fold between different genes (e.g., Table 4.1 of Graur and Li, 2000).

Supplemental Table 4a–c also show that, apart from chloroplast *matK* genes in which nonsynonymous sites evolve about 10 times faster than that of other chloroplast genes, $K_a$ values for different genes within the same genome are very similar and do not vary by more than two- to four-fold. For example, the average $K_a$ values for all angiosperms vary from 0.014 to 0.026 for chloroplast genes (excluding *matK* genes), from 0.013 to 0.054 for mitochondrial genes and from 0.043 to 0.116 for nuclear genes (Supplemental Table 4a–c). This is very different from what has been observed in the nuclear protein coding genes of mammals where nonsynonymous rates have been observed to vary by more than 300-fold between different genes (e.g., Table 4.1 of Graur and Li, 2000). The similar nonsynonymous rates we observe between different genes within each genome suggest that these different genes evolve under similar selective constrains within each genome.

Table 3 shows the $K_s$ and $K_a$ values of the concatenated genes of each genome for the eight seed plant groups. This table also shows which values are significantly different between the three different genomes.

Table 4 shows the $K_s$ ratios of the concatenated genes of each genome for the eight seed plant groups. These values are the ratios obtained by dividing the $K_s$ or $K_a$ values shown in Table 3 by their respective mitochondrial $K_s$ or $K_a$ values. These results show that the differences in relative synonymous rates between the three genomes are smaller within gymnosperms (mitochondrial:chloroplast:nuclear ratios of 1:2:4) than within angiosperms (mitochondrial:chloroplast:nuclear ratios of 1:3:16). On the other hand, these ratios are similar within both angiosperms and gymnosperms. The chloroplast ratios of monocots, eudicots and basal angiosperms are 3.7, 2.7 and 3.3, respectively, and the nuclear ratios of the same groups are 17.2, 15.5 and 19.6, respectively. The chloroplast ratios of gnetales and other gymnosperms are both 1.9 and the nuclear ratios of the same groups are 5.3 and 4.3, respectively.

## 4. Discussion

Our results show that the mitochondrial genes of all seed plant groups evolve more slowly than their chloroplast genes and the chloroplast genes of all seed plant groups evolve more slowly than their nuclear genes (Tables 3 and 4). In angiosperms, we obtained mitochondrial:chloroplast:nuclear $K_s$ ratios of 1:3:10 with the method of Li et al. (Table 4). This is very similar to the 1:3:12 ratios previously obtained by Wolfe and colleagues who had also used the Li et al. method to calculate their ratios (Wolfe et al., 1987, 1989). This suggest that the divergence dates they used to calculate their substitution rates were reasonable and that the variability introduced by comparing different genes between different species did not overly influence the ratios they obtained.

The main difference between our results and those of Wolfe et al. (1987, 1989) is that the ML method gives higher nuclear $K_s$ ratios. In angiosperms, the ML mitochondrial:chloroplast:nuclear $K_s$ ratios are 1:3:16 instead of 1:3:10. This not unexpected because, as mentioned above, the ML method is better at correcting for multiple hits between divergent sequences. This is reflected by the fact that the largest $K_s$ distance calculated by the Li et al. method on our nuclear data set was 2.29 (between *T. canadensis* and *A. thaliana*) whereas the corresponding $K_s$ distance calculated by the ML method was 5.84 (results not shown).

Our results show that the relative synonymous rates between the three genomes are different between gymnosperms (mitochondrial:chloroplast:nuclear ratios of 1:2:4) and angiosperms (mitochondrial:chloroplast:nuclear ratios of 1:3:16) but that these ratios are similar within both angiosperms and gymnosperms (Table 4). These observations suggest that the factors responsible for the difference in mitochondrial:chloroplast:nuclear ratios between angiosperms and gymnosperms are common to all members of each of these two seed plant groups.

Although the mitochondrial:chloroplast:nuclear ratios are very similar within both angiosperms and gymnosperms, there is nevertheless some variation within each of these groups (Table 4). For example, in angiosperms, the mitochondrial:chloroplast:nuclear ratios of eudicots (1:2.7:15.5) are lower than that of monocots (1:3.7:17.2). Similarly, the mitochondrial:nuclear ratios of other gymnosperms (1:4.3) is slightly lower than that of gnetales (1:5.3). This suggest that the factors responsible for the difference in mitochondrial:chloroplast:nuclear ratios between angiosperms and gymnosperm are also somewhat variable within each of these two taxonomic groups.

Our results show that rates of synonymous substitution of the mitochondrial genes of the plant species we analyzed are low and similar between different genes (Table 3, Supplemental Table 4b). However, recent studies have demonstrated that exceptions do exist and that the rate of synonymous substitution in the mitochondrial genomes of some plant genera can be up to ten thousand time higher than that of other taxa (Cho et al., 2004; Parkinson et al., 2005; Mower et al., 2007).

The rates of synonymous and nonsynonymous substitutions are not independent within each genome. In fact, the $K_s$ and $K_a$ values

**Table 3**
Snonymous and nonsynonymous substitutions of concatenated genes

| Genes | $K_s$—LWL | $K_a$—LWL | $K_s$—ML | $K_a$—ML |
|---|---|---|---|---|
| *Mitochondrial* | | | | |
| Monocots | 0.096 ± 0.028 | 0.011 ± 0.000 | 0.093 ± 0.025 | 0.010 ± 0.004 |
| Eudicots | 0.163 ± 0.012 | 0.021 ± 0.000 | 0.160 ± 0.011 | 0.019 ± 0.002 |
| Basal angiosperms | 0.067 ± 0.007 | 0.014 ± 0.000 | 0.066 ± 0.006 | 0.013 ± 0.001 |
| All angiosperms | 0.129 ± 0.005 | 0.020 ± 0.000 | 0.128 ± 0.005 | 0.018 ± 0.001 |
| Gnetales | 0.303 ± 0.058 | 0.064 ± 0.001 | 0.323 ± 0.068 | 0.051 ± 0.009 |
| Other gymnosperms | 0.202 ± 0.022 | 0.059 ± 0.001 | 0.194 ± 0.022 | 0.058 ± 0.005 |
| All gymnosperms | 0.282 ± 0.020 | 0.070 ± 0.001 | 0.282 ± 0.021 | 0.065 ± 0.004 |
| All seed plants | 0.233 ± 0.007 | 0.046 ± 0.001 | 0.233 ± 0.007 | 0.042 ± 0.001 |
| *Chloroplast* | | | | |
| Monocots | 0.360 ± 0.100 | 0.039 ± 0.011* | 0.346 ± 0.099 | 0.039 ± 0.010 |
| Eudicots | 0.452 ± 0.014*** | 0.054 ± 0.003** | 0.439 ± 0.014** | 0.052 ± 0.002** |
| Basal angiosperms | 0.229 ± 0.013* | 0.032 ± 0.001 | 0.218 ± 0.012* | 0.031 ± 0.001* |
| All angiosperms | 0.402 ± 0.012* | 0.049 ± 0.001* | 0.388 ± 0.012* | 0.047 ± 0.001* |
| Gnetales | 0.625 ± 0.055* | 0.093 ± 0.010 | 0.615 ± 0.058* | 0.088 ± 0.009 |
| Other gymnosperms | 0.394 ± 0.018 | 0.054 ± 0.002 | 0.376 ± 0.018 | 0.053 ± 0.002 |
| All gymnosperms | 0.603 ± 0.033 | 0.091 ± 0.005 | 0.605 ± 0.035 | 0.087 ± 0.005 |
| All seed plants | 0.618 ± 0.013 | 0.084 ± 0.002 | 0.613 ± 0.014 | 0.082 ± 0.002 |
| *Nuclear* | | | | |
| Monocots | 1.121 ± 0.258/* | 0.058 ± 0.009/* | 1.595 ± 0.493/* | 0.039 ± 0.004/** |
| Eudicots | 1.473 ± 0.049***/*** | 0.090 ± 0.005*/*** | 2.476 ± 0.135***/*** | 0.066 ± 0.004/** |
| Basal angiosperms | 0.944 ± 0.052*/** | 0.043 ± 0.002/* | 1.295 ± 0.082**/** | 0.029 ± 0.002 |
| All angiosperms | 1.288 ± 0.030**/*** | 0.068 ± 0.002/* | 2.107 ± 0.085*/* | 0.046 ± 0.001 |
| Gnetales[a] | 1.216 | 0.072 | 1.719 | 0.054 |
| Other gymnosperms[a] | 0.720 ± 0.034*/** | 0.037 ± 0.003 | 0.837 ± 0.046*/** | 0.029 ± 0.003*/ |
| All gymnosperms[a] | 0.944 ± 0.050/* | 0.053 ± 0.004 | 1.234 ± 0.086/* | 0.041 ± 0.003 |
| All seed plants | 1.388 ± 0.019*/** | 0.076 ± 0.001 | 2.441 ± 0.062/* | 0.053 ± 0.001 |

*Notes.* Values are substitutions/site ± standard error. $K_s$—LWL, $K_s$ measured using the method of Li et al.; $K_a$—LWL, $K_s$ measured using the method of Li et al.; $K_s$—ML, $K_s$ measured using the ML method; $K_a$—ML, $K_a$ measured using the ML method. [a]Based on *rpb1* and *rpb2* sequences only. Significance levels for chloroplast sequences are compared to corresponding mitochondrial sequences. Significance levels for nuclear sequences are compared to corresponding chloroplast (before the "/" symbol) and mitochondrial sequences (after the "/" symbol).

* $P < 0.05$.
** $P < 0.01$.
*** $P < 0.001$.

**Table 4**
$K_s$ and $K_a$ ratios of concatenated genes

| Genes | $K_s$—LWL | $K_a$—LWL | $K_s$—ML | $K_a$—ML |
|---|---|---|---|---|
| *Mitochondrial* | | | | |
| Monocots | 1 | 1 | 1 | 1 |
| Eudicots | 1 | 1 | 1 | 1 |
| Basal angiosperms | 1 | 1 | 1 | 1 |
| All angiosperms | 1 | 1 | 1 | 1 |
| Gnetales | 1 | 1 | 1 | 1 |
| Other gymnosperms | 1 | 1 | 1 | 1 |
| All gymnosperms | 1 | 1 | 1 | 1 |
| All seed plants | 1 | 1 | 1 | 1 |
| *Chloroplast* | | | | |
| Monocots | 3.8 | 3.5 | 3.7 | 3.9 |
| Eudicots | 2.8 | 2.6 | 2.7 | 2.7 |
| Basal angiosperms | 3.4 | 2.3 | 3.3 | 2.4 |
| All angiosperms | 3.1 | 2.5 | 3.0 | 2.6 |
| Gnetales | 2.1 | 1.5 | 1.9 | 1.7 |
| Other gymnosperms | 2.0 | 0.9 | 1.9 | 0.9 |
| All gymnosperms | 2.1 | 1.3 | 2.1 | 1.3 |
| All seed plants | 2.7 | 1.8 | 2.6 | 2.0 |
| *Nuclear* | | | | |
| Monocots | 11.7 | 5.3 | 17.2 | 3.9 |
| Eudicots | 9.0 | 4.3 | 15.5 | 3.5 |
| Basal angiosperms | 14.1 | 3.1 | 19.6 | 2.2 |
| All angiosperms | 10.0 | 3.4 | 16.5 | 2.6 |
| Gnetales | 4.0 | 1.1 | 5.3 | 1.1 |
| Other gymnosperms | 3.6 | 0.6 | 4.3 | 0.5 |
| All gymnosperms | 3.3 | 0.8 | 4.4 | 0.6 |
| All seed plants | 6.0 | 1.7 | 10.5 | 1.3 |

*Notes.* $K_s$—LWL, $K_s$ measured using the method of Li et al.; $K_a$—LWL, $K_s$ measured using the method of Li et al.; $K_s$—ML, $K_s$ measured using the ML method; $K_a$—ML, $K_a$ measured using the ML method.

shown in Table 4 are strongly correlated within each genome (Pearson's correlation coefficients of 0.86, 0.98 and 0.86 for the mitochondrial, chloroplast and nuclear $K_s$ and $K_a$ values, respectively). Yet these within genome correlations do not influence the between genome variation in $K_s$ values we observed. For example, whereas the average $K_a$ value of monocots is 0.039 for both chloroplast and nuclear genes, the average monocot $K_s$ value of nuclear genes (1.595) is almost five times that of chloroplast genes (0.346; Table 4).

Several factors (hypotheses) have been suggested to explain the difference in evolutionary rates observed in diverse organisms. These include differences in generation times, metabolic rates, fidelity of the replication process, DNA repair efficiency and intensity of purifying selection (Graur and Li, 2000). The diversity of species included in each of our angiosperm and gymnosperm data sets makes it unlikely that differences in generation times, metabolic rates and intensity of purifying selection could account for the overall difference observed between these two groups. It seems more likely that differences in replication fidelity or DNA repair efficiency, brought about by changes in the coding sequences of genes coding for enzymes involved in these two processes, are responsible for the differences observed. Since male-driven evolution of mitochondrial and chloroplast sequences has been shown to occur in gymnosperms, and that different gymnosperm species inherit their mitochondria and/or chloroplast from different parents, it might also be that some of the different gymnosperm mitochondrial:chloroplast:nuclear ratios reflect differences in parental inheritance of mitochondrial and chloroplast sequences in different gymnosperm species (Mogensen, 1996; Whittle and Johnston, 2002). Another possibility, kindly suggested by a reviewer, is that increased $K_s$ nuclear values might be caused by different species

using different tRNA pools to translate their nuclear genes. For example, increased $K_s$ values will be observed within a taxonomic group if some of its species have many tRNAs translating the GCR codons of alanine (where R is either an A or a G) and few tRNAs translating the GCY codons of this amino acid (where Y is either an C or a T), whereas the abundance of these two tRNA groups is reversed in other species, because these differences will select for GCR codons in some species and GCY codons in other species. The differences in mitochondrial:chloroplast:nuclear ratios also indicate that the nucleotide substitution rate of one or more of the organelle's genomes has changed. Solving this issue will require reliable divergence times to calculate the absolute evolutionary rates of mitochondrial, chloroplast and nuclear genes from these diverse seed plant species.

In conclusion, our results show that the mitochondrial, chloroplast and nuclear genomes of seed plant groups have different synonymous substitutions rates, that these rates are different in different seed plant groups and that gymnosperms have smaller ratios than angiosperms. Although these rates and ratios are expected to change with different taxonomic sampling and with the methods used to measure them, our results clearly show that the synonymous substitutions rates of mitochondrial, chloroplast and nuclear genomes are different in different seed plant groups. It will be interesting to try to find the reason(s) responsible for these synonymous substitution rate variations.

## Acknowledgments

We thank David Benovoy, Robert T. Morris and Robert Carter (University of Ottawa) for their help and comments during this work. We thank both reviewers for their useful and constructive comments. We thank Robert Carter for writing us a script to extract LWL85 data from the output files of the YN00 program. We also thank Hugette Allard (University of Ottawa), Chris Graham (Royal Botanical Gardens, Hamilton, Ont.), Richard Olmstead (Washington University), Yin-Long Qiu (University of Massachusetts) and Alan Whittemore (US National Arboretum) for providing us with fresh plants or plant leaves for mRNA extractions. This work was supported by an OGS scholarship to J. X. and a Discovery Grant from the National Science and Engineering Research Council of Canada to G. D.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2008.09.009.

## References

Brown, W.M., George Jr., M., Wilson, A.C., 1979. Rapid evolution of animal mitochondrial DNA. Proc. Natl. Acad. Sci. USA 76, 1967–1971.

Brown, W.M., Prager, E.M., Wang, A., Wilson, A.C., 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution. J. Mol. Evol. 18, 225–239.

Burleigh, J.G., Mathews, S., 2004. Phylogenetic signal in nucleotide data from seed plants: implications for resolving the seed plant tree of life. Am. J. Bot. 91, 1599–1613.

Cho, Y., Mower, J.P., Qiu, Y.L., Palmer, J.D., 2004. Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. Proc. Natl. Acad. Sci. USA 101, 17741–17746.

Easteal, S., Collet, C.C., Betty, D.J., 1995. The Mammalian Molecular Clock. Springer & Landes, Austin, TX.

Graur, D., Li, W.-H., 2000. Fundamentals of Molecular Evolution, Second ed. Sinauer Associates, Sunderland, Massachusetts.

Hajibabaei, M., Xia, J., Drouin, G., 2006. Seed plant phylogeny: gnetophytes are derived conifers and a sister group to Pinaceae. Mol. Phylogenet. Evol. 40, 208–217.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.

Jansen, R.K., Cai, Z., Raubeson, L.A., Daniell, H., Depamphilis, C.W., Leebens-Mack, J., Müller, K.F., Guisinger-Bellian, M., Haberle, R.C., Hansen, A.K., Chumley, T.W., Lee, S.B., Peery, R., McNeal, J.R., Kuehl, J.V., Boore, J.L., 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc. Natl. Acad. Sci. USA 104, 19369–19374.

Kuzoff, R.K., Gasser, C.S., 2000. Recent progress in reconstructing angiosperm phylogeny. Trends Plant Sci. 5, 330–336.

Li, W.-H., Wu, C.I., Luo, C.C., 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. 2, 150–174.

Luo, J., Yoshikawa, N., Hodson, M.C., Hall, B.D., 2007. Duplication and paralog sorting of RPB2 and RPB1 genes in core eudicots. Mol. Phylogenet. Evol. 44, 850–862.

Magallon, S.A., Sanderson, M.J., 2005. Angiosperm divergence times: the effect of genes, codon positions and time constraints. Evolution 59, 1653–1670.

Mathews, S., Donoghue, M.J., 1999. The root of angiosperm phylogeny inferred from duplicate phytochrome genes. Science 286, 947–950.

Miyata, T., Hayashida, H., Kikuno, R., Hasegawa, M., Kobayashi, M., Koike, K., 1982. Molecular clock of silent substitution: at least six-fold preponderance of silent changes in mitochondrial genes over those in nuclear genes. J. Mol. Evol. 19, 28–35.

Mogensen, H.L., 1996. The hows and whys of cytoplasmic inheritance in seed plants. Am. J. Bot. 83, 383–404.

Moore, M.J., Bell, C.D., Soltis, P.S., Soltis, D.E., 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proc. Natl. Acad. Sci. USA 104, 19363–19368.

Mower, J.P., Touzet, P., Gummow, J.S., Delph, L.F., Palmer, J.D., 2007. Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. BMC Evol. Biol. 7, 135.

Muse, S.V., 1996. Estimating synonymous and nonsynonymous substitution rates. Mol. Biol. Evol. 13, 105–114.

Nickerson, J., Drouin, G., 2004. The sequence of the largest subunit of RNA polymerase II is a useful marker for inferring seed plant phylogeny. Mol. Phylogenet. Evol. 31, 403–415.

Oxelman, B., Yoshikawa, N., McConaughy, B.L., Luo, J., Denton, A.L., Hall, B.D., 2004. RPB2 gene phylogeny in flowering plants, with particular emphasis on asterids. Mol. Phylogenet. Evol. 32, 462–479.

Parkinson, C.L., Mower, J.P., Qiu, Y.L., Shirk, A.J., Song, K., Young, N.D., DePamphilis, C.W., Palmer, J.D., 2005. Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae. BMC Evol. Biol. 5, 73.

Soltis, P.S., Soltis, D.E., Savolainen, V., Crane, P.R., Barraclough, T.G., 2002. Rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence for molecular living fossils. Proc. Natl. Acad. Sci. USA 99, 4430–4435.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Whittle, C.A., Johnston, M.O., 2002. Male-driven evolution of mitochondrial and chloroplastidial DNA sequences in plants. Mol. Biol. Evol. 19, 938–949.

Wolfe, K.H., Li, W.-H., Sharp, P.M., 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast and nuclear DNAs. Proc. Natl. Acad. Sci. USA 84, 9054–9058.

Wolfe, K.H., Sharp, P.M., Li, W.-H., 1989. Rates of synonymous substitutions in plant nuclear genes. J. Mol. Evol. 29, 208–211.

Yang, Z., Nielsen, R., 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol. Biol. Evol. 1, 32–43.

Yang, Z., 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.

Zuckerkandl, E., Pauling, L., 1965. Evolutionary divergence and convergence in proteins. In: Bryson, V., Vogel, H.J. (Eds.), Evolving Genes and Proteins. Academic Press, New York, pp. 97–166.