東海大學生命科學系
碩士論文

指導教授：劉少倫 博士

Dr. Shao-Lun Liu

四個幽蟌(*Euphaea*)姊妹種粒線體基因作為族群遺傳及親緣關係分子標記的效能探討

Population Genetic and Phylogenetic Utility of Mitochondrial Genes of Four Sibling *Euphaea* Damselflies

研究生：鄭筠潔

Yun-Chieh Cheng

中華民國一百零五年八月

東海大學生命科學系碩士論文

四個幽蟌(*Euphaea*)姊妹種粒線體基因作為族群遺傳及親緣關係分子標記的效能探討

Population Genetic and Phylogenetic Utility of Mitochondrial Genes of Four Sibling *Euphaea* Damselflies

研究生：鄭筠潔

Yun-Chieh Cheng

指導教授：劉少倫 博士

Dr. Shao-Lun Liu

中華民國一百零五年八月

東海大學生命科學系
碩士論文學位考試審定書

生命科學系碩士班研究生 鄭筠潔 君所撰寫之論文

（中文）

四個幽蟌(Euphaea)姊妹種粒線體基因作為族群遺傳及親緣關係分子標記的效能探討

（英文）

Population Genetic and Phylogenetic Utility of Mitochondrial Genes of Four Sibling Euphaea Damselflies

經本委員會審定通過，特此證明。

學位考試委員會　召集人 劉少倫 （簽名）

委　員 林仲平

李壽先

中華民國 105 年 6 月 20 日

致謝

這篇論文的完成，首先要感謝的是林仲平老師，非常感謝老師對實驗分析及論文撰寫上的細心指導與幫助。也謝謝特地前來幫忙口試的委員們，師大的李壽先老師及東海的劉少倫老師，謝謝您們的指教與建議。在分析上，謝謝若凡學姊及祐薰學姊及師大的秉宏學長不厭其煩的幫助我解決在研究分析上遇到的困難。另外，提供豆娘標本及 DNA 序列的實驗室學長們，黃仁磐學長及李逸鴻學長，謝謝你們。

從大學時期以來陪伴我的實驗室學長姐學弟妹們，謝謝你們教會我做分生實驗、幫我一起出野外並指導我如何飼養各式昆蟲以及分析各種數據，並且在各種預講時給予意見。當然還有那些餘興活動，感謝大家的大力支持與參與。就讀碩士班的這段生涯，一路走來，總是受到許多人的幫助扶持，不論是在東海或是師大或是校外的好朋友，真的很高興能夠認識你們，謝謝你們這些日子裡帶給我的溫暖時光，我會懷念的。最後，一直在旁邊默默陪伴的家人們，真的很感激你們對我的支持。帶著所有人的祝福，我將勇敢朝下一個階段邁進，謝謝大家！

Contents

Abstract (In Chinese)

一直以來，經由母系遺傳的粒線體基因被認為是研究族群遺傳及系統發生學的有效分子標記。但近期研究卻發現粒線體基因或許並不適合此用途，原因有二：一、由於其經歷過頻繁發生的正向選汰，此時有利的突變及其相鄰的基因會被固定下來(genetic draft)，這種狀況的發生使得粒線體基因擁有比中性理論預期還低的遺傳多樣性。二、粒線體基因過快的演化速率使各分子間的親緣關係難以表現出來。但，這些研究通常都以較高階的分類群為主，粒線體基因在種間的效能表現如何仍有許多未知。本研究透過不同族群大小的兩對姊妹群幽螽屬豆娘─*E. formosa* + *E. yayeyamana*；*E. decorata* + *E. ornata*─的 8 個核基因、13 個粒線體基因來對選汰壓力及基因效能進行分析。結果顯示，粒線體基因較核基因受到更強烈的負向選汰影響，而負向選汰是影響這些豆娘粒線體基因及核基因演化的主因。粒線體基因無論是在成對距離的分化度、α 值、簡約訊息位點、不變位點的比例及相對置換率這些基因效能參數上都比核基因還要為高。在這些參數的平均中，粒線體基因的非蛋白質轉譯區比蛋白質轉譯區擁有較高的解析度。另外，核基因的外顯子提供的資訊較內插子要少。在相關性分析上，*A+T-rich* 及 *nad2* 這兩個粒線體基因表現得最好，並能提供最多的資訊。研究顯示粒線體基因適合用作族群遺傳及系統發生學的分析之用，但需考慮如何選擇基因。

關鍵字：粒線體基因、有效族群大小、正向選汰、負向選汰、基因效能、蜻蛉目

Abstract

The maternal inherited mitochondrial genes have been considered as efficient genetic markers for population genetics and phylogenetic studies. However, recent studies suggested that mitochondrial genes might not be suitable for these purposes because (a) these genes have lower genetic diversity than expected from neutral evolution due to fixation of beneficial mutations by frequent positive selection (i.e. genetic draft), and (b) these genes evolved too rapidly to represent phylogenetic relationships. Nevertheless, these studies primarily focused on analyzing higher taxa, leaving the utilities of mitochondrial genes on species level analyses largely unknown. In this study, we investigated the pattern of selection and phylogenetic utilities of eight nuclear genes and 13 mitochondrial genes in two pairs of sibling *Euphaea* damselflies (*E. formosa* of Taiwan + *E. yayeyamana* of Yaeyama Islands; *E. decorata* of SE Asia + *E. ornata* of Hainan Islands) with various population sizes. Our results showed that, comparing to nuclear genes, mitochondrial genes experienced stronger negative selection, which was the dominance force in the evolution of both mitochondrial and nuclear genes in these damselflies. Mitochondrial genes had higher rates than nuclear genes in the pairwise distance divergence, α (shape of gamma distribution), PI (parsimony informative site), Pi (proportion of invariant sites), and the relative substitution rate. Among the averages of parameters, the mitochondrial

protein non-coding regions had higher resolution than mitochondrial protein-coding region. In addition, the exons of nuclear genes provided less information than the introns. Correlation analyses demonstrated that mitochondrial *A+T-rich* and *nad2* genes performed better and were more informative. In conclusion, our results showed that mitochondrial genes are suitable for genetics and phylogenetic analyses, but the selection of genes need to be considered.

Keywords: Mitochondrial genome, effective population size, genetic draft, negative selection, phylogenetic utility, Odonata

Introduction

Population genetic and phylogenetic studies employ a range of genetic markers, including allozymes, microsatellites, single nucleotide polymorphism (SNP), mitochondrial, and nuclear genes that differ in their utility and resolution from populations, species, to higher taxonomic groups (Galtier *et al*., 2009; Dupuis *et al*., 2012). Among these markers, the maternally inherited mitochondrial genes of animals is popular because of their ease of amplification, large copy numbers, rapid mutation rates, and non-recombination which makes them efficient and effective choices for inferring population genetic structures and reconstructing phylogenies (Sunnucks, 2000; Simon *et al*., 1994; Simon *et al*., 2006).

The effective population size (N_e) is the number of individuals who can exchanges its genetic factors in an ideal population (Wright 1931), which is one of the important parameters in population genetics and has been used to effectively assess the impacts of overfishing (Hauser *et al*., 2002; Roman & Palumbi, 2003) or tracking historical population dynamics (Miller & Waits, 2003; Luikart *et al*., 2010). Theoretically, N_e is related to the level of genetic diversity in a population. Earlier empirical studies have showed that the N_e of a given population is positively correlated with its genetic diversity (Frankham *et al*., 1996; Mulligan *et al*., 2006; Heller *et al*., 2010; Luikart *et al*., 2010). In general, the population size of an

organism determine the relative strength of genetic drift and natural selection acting on allele frequencies of the population (Nielsen & Slatkin, 2013). The stochastic event of genetic drift has a stronger effect on genetic diversity of small populations, whereas natural selection is the dominant force in large populations (Meiklejohn *et al*., 2007).

Despite the popularity of mitochondrial DNA in ecology and evolutionary studies, earlier studies suggested that the genetic diversity of mitochondrial DNA are not positively correlated with the population sizes (Galtier *et al*., 2009). Bazin and colleagues showed that the mammals have higher mitochondrial genetic diversity than that of the insects and mollusks, which in general have much larger population sizes (Bazin *et al*., 2006; Piganeau & Eyre-Walker, 2009). One hypothesized mechanism for the decreased mitochondrial diversity in large populations is through genetic draft, a process which fixes advantageous mutations and reduces mitochondrial diversity by frequent positive selection. Because the effect of genetic draft is much stronger in larger populations (Halliburton, 2004; Bazin *et al*., 2006; Meiklejohn *et al*., 2007), invertebrates with larger population sizes (such as insects, mollusks) therefore tend to have lower mitochondrial genetic diversity than that of vertebrates with smaller populations. Due to maternal inheritance, mitochondrial genes have a smaller population size than nuclear genes. Previous studies suggested that mitochondrial genes cannot be used as reliable genetic markers for estimating effective population

size (Bazin *et al*., 2006; Galtier *et al*., 2009). However, the level and frequency of genetic draft acting on mitochondrial genomes is largely unknown for most species (Roux *et al*., 2014).

Recent studies also suggested that mitochondrial genes evolve too rapidly and show nucleotide substitution patterns that are problematic for higher level phylogenetics (Lin & Danforth, 2004; Danforth *et al*., 2006). Comparing to nuclear genes, mitochondrial genes showed higher inconsistency among phylogenetic signals (lower values of CI, consistency index) and less symmetrical transformation rate matrices of nucleotide substitution, suggesting that they are lower quality genes for phylogenetic analysis (Lin & Danforth, 2004; Danforth *et al*., 2006). These studies primarily analyzed deeper divergence (Mesozoic and older) of higher level (generic, tribal) phylogenetic relationships and evaluated only the most commonly used mitochondrial genes such as cytochrome c oxidase subunit 1 (*cox1*) and cytochrome c oxidase subunit 2 (*cox2*). The relative phylogenetic utility among mitochondrial genes and the utility of mitochondrial genes at shallow divergence (Pleistocene and younger) of lower level (species, population) relationships are little known, especially for less frequently utilized mitochondrial genes.

In this study, we investigated the pattern of selection and nucleotide substitution between mitochondrial and nuclear genes at species and population level using

112 *Euphaea* damselflies (Insecta, Odonata, Euphaeidae) as an example. We sequenced
113 complete mitochondrial genomes of three closely related *Euphaea* species (*E.*
114 *decorata*, *E. ornata* and *E. yayeyamana*) and obtained mitochondrial genome
115 sequences of sibling species, *E. formosa* from GenBank (Lin *et al*., 2010). We then
116 used 13 protein-coding genes and eight nuclear genes (Lee & Lin, 2012) of these four
117 species to estimate the level of positive selection (genetic draft) and evaluated the
118 phylogenetic utility of these genes. The geographic distribution of the four species
119 indicate that *E. yayeyamana* of the Yaeyama islands has the smallest population
120 among the four *Euphaea* species. *Euphaea formosa* and *E. ornata* of Taiwan and
121 Hainan Island, respectively, represent species with a medium population. *Euphaea*
122 *decorata* is a widespread species of South-East Asia, distributing from Fujian,
123 Guangdong and Guanhxi (China) to Vietnam, representing a species with the largest
124 population size among the four species. First, we predict that the species with the
125 largest population size, *E. decorata*, experienced the strongest selective constrain than
126 the two species with medium population size. The species with the smallest
127 population size, *E. yayeyamana*, has the lowest level of positive selection. Comparing
128 mitochondrial with nuclear genes, the nuclear genes are predicted to exhibit a higher
129 level of positive selection than that of the mitochondrial genes because the effective
130 population size of nuclear genes are in general four times larger than mitochondrial

genes (Ballard & Whitlock, 2004). Secondly, we compared the phylogenetic utility among all mitochondrial genes and between mitochondrial and nuclear genes by calculating the parameters of nucleotide substitution and inferring patterns of rate variation among genes. Specifically, we examined the differences in substitution rate among mitochondrial and nuclear genes, and tested a positive relationships between base composition (A+T bias) and relative rates, between CI (consistency index) and α (the shape parameter of gamma distribution), CI and relative rate, PI sites (parsimony informative sites) and α, PI sites and relative rate and between Pi (proportion of invariant sites) and α. The comparisons help us to see the relationships between characteristics of each gene, and to know which gene is the most diverse.

Materials and Methods

Mitochondrial Genome Sequencing

The damselfly specimens of *E. decorata*, *E. ornata*, and *E. yayeyamana* used for genome sequencing were collected in Hong Kong, Hainan Island, and Yaeyama Islands, respectively (Huang & Lin, 2011; Lee & Lin, 2012). The specimens were placed in 95% EtOH after capture and then stored in a -80℃ freezer upon returning to the laboratory. Total genomic DNAs were extracted using the standard protocol (Yamauchi *et al*., 2004) and then preserved in the -20℃ freezer. Two sets of long PCR primers (L2020-CO1/H12230-16S; L12167-16S/H5244-CO3) were used to amplify

151 the complete mitochondrial genome as two overlapping DNA fragments following the
152 recommended thermal cycles (Yamauchi *et al*., 2004). The DNA sequencing was then
153 performed by primer walking procedure with the 22 published primers (Lin *et al*.,
154 2010). The genome sequences were edited and assembled following the established
155 procedure (Lin *et al*., 2010), and annotated with reference to the mitochondrial
156 genome of *E*. *formosa* (GenBank, accession number: HM126547). The three complete
157 mitochondrial genomes were submitted to GenBank (accession numbers: *E*.
158 *yayeyamana*, KF718293; *E*. *decorata*, KF718294; *E*. *ornata*, KF718295). A total of
159 75 mitochondrial (*cox2*, NADH dehydrogenase subunit 5 (*nad5*)) and 136 nuclear
160 gene sequences (actin (*act*), arrestin2 (*arr*), abnormal wing discs 2 (*awd2*), elongation
161 factor 1alpha (*EF1α*), *ferritin (fer*), myosin light chain (*mlc*), *opsin*, succinate
162 dehydrogenase B (*sdhB*)) was obtained from NCBI (Appendix 1). Additional one
163 *cox2* (EfAa12wi), eight *nad5* (EfAa1, EfAa2, EfBa1, EfBa5, EfBa12, EfDa4, EfTa3,
164 TW1) and one *EF1α* (EfAa9) gene sequences of *E*. *formosa* were sequenced for
165 calculating d_N/d_S ratio at population level. The chromatograms of the DNA sequences
166 were manually edited in MEGA 5.2 (Tamura *et al*., 2011). The nucleotide sequences
167 of protein coding genes were translated into amino acid sequences using invertebrate
168 mitochondrial genetic codes in the SIB ExPASy Bioformatics Resources Portal
169 (Gasteiger *et al*., 2003) to confirm reading frames and sequencing errors. The

haplotype numbers of each genes were inferred using dnaSP v5 (Librado & Rozas, 2009). For population-level analysis, we analyzed 137 nuclear and 84 mitochondrial sequences of 10 genes (mitochondrial: *cox2*, *nad5*; nuclear: *act*, *arr*, *awd2*, *EF1α*, *fer*, *mlc*, *opsin*, *sdhB*). For species-level analysis, we used the 13 mitochondrial protein-coding and non-coding genes for four *Euphaea* species to compare with eight nuclear genes.

Nonsynonymous/synonymous substitution ratio (d_N/d_S, ω)

The d_N/d_S ratio (ω) of the obtained DNA sequences was calculated to test for mitochondrial signature of natural selection (Miyata & Yasunaga, 1980; Yang & Bielawski, 2000). A value of ω significantly greater than 1 indicates positive selection; whereas a value of ω significantly less than 1 suggests negative selection (Yang & Bielawski, 2000). The extremely large and small values of ω (>100, <0.001) were excluded in the analysis because these extreme values were mainly due to the low number of substitution. The nucleotide substitution model was selected in MEGA 5.2 using BIC (Bayesian Information Criterion). The models selected for individual genes were as followings: HKY: *act*, *arr*, *atp6*, *atp8*, *awd2*, *cox2*, *nad3*, *nad4l*, *nad6*, *EF1α*, *fer*, *mlc*, *opsin* and *sdhB*; HKY+G: *nad2*, *nad4*, *nad5* and *cob*; TN93+G: *cox1* and *nad1*; TN93: *cox3*. The individual gene trees of four *Euphaea* species was reconstructed in MEGA 5.2 using maximum likelihood. The ω of mitochondrial

189 protein-coding and nuclear genes were calculated using the branch model in codeml
190 option of PAML v.4.7 (Yang, 2007). Two models were used to compare the strength
191 of selection using maximum likelihood: model= 0, represents only one ω ratio for all
192 four species ($\omega_{T,}$ which refers to total ω); model= 2, assumes each species has
193 different ω ratios (ω_d, ω_o, ω_f and ω_y, which refers to the ω of *E. decorata*, *E. ornata*,
194 E. *formosa* and *E. yayeyamana*). The strength of selection are different among species
195 if the null hypothesis (model= 0) was rejected. The Hyphy program in MEGA 6.0
196 (Tamura *et al*., 2013) was used to calculate the level of selection across codon sites
197 within each genes. The results were illustrated using a sliding window plot.
198 *Phylogenetic analysis and nucleotide substitution pattern*
199 We used 24 genes (13 protein-coding and three non-coding mitochondrial genes,
200 and eight nuclear genes) to co-estimate the species tree of four *Euphaea* species in
201 *BEAST v.1.8.2 (Drummond *et al*., 2012) (Appendix 2). The nucleotide substitution
202 models for each genes were the same as in likelihood analysis of PAML, with
203 additional models for two ribosome genes (HKY: small subunit 18S rRNA (*s-rRNA*)
204 and *l-rRNA*) and *A+T-rich* (HKY+G). The Markov chain Monte Carlo (MCMC)
205 procedure was executed for 1×10^8 generations with trees sampled for every 1×10^4
206 generations. The initial 2×10^7 generations were discarded as burn-in. The
207 convergence of MCMC runs were assessed using values of the effective sample size

208 (ESS >741) in Tracer v.1.6 (Rambaut *et al*., 2014).

209 The pairwise genetic distance of the haplotypes of four *Euphaea* species (Table 3)
210 were calculated in MEGA 6.0 (Tamura *et al*., 2013). The number of substitutions per
211 site within and between species was calculated using Tajima-Nei model (Tajima &
212 Nei, 1984). We corrected the pairwise distance using the mutation rate of 1.77×10^{-8}
213 mutation /site/year of *cox1* (Papadopoulou *et al*., 2010) for all mitochondrial genes. A
214 substitution rate of 1.2×10^{-8} mutation /site/year of nuclear *arr* (Moriyama &
215 Gojobori, 1992) was used for nuclear genes. The PI ratio (parsimony informative sites
216 per codon) were calculated in MEGA 6.0.

217 Two nucleotide substitution models, GTR+SSR (site-specific rates) and
218 GTR+I+G (invariant sites and gamma distribution), were used to calculate the relative
219 substitution rates among genes in MrBayes v.3.2 (Huelsenbeck & Ronquist, 2001).
220 The gene partitions included 13 mitochondrial protein-coding genes, three non-coding
221 regions (*s-rRNA*, *l-rRNA* and *A+T-rich*) and eight nuclear genes. For site-specific
222 rates (SSR), we partitioned the substitution rates of a given protein-coding gene into
223 three codon sites. For GTR+SSR analyses, the MCMC was run for 1×10^{7}
224 generations with trees sampled for every 1,000 generations. In GTR+I+G model, we
225 obtained the shape parameter (α of gamma distribution) and Pi (the proportion the
226 invariant sites) to estimate the level of among site rate heterogeneity and Q matrix,

transformation rate matrix. The α value estimate the rate heterogeneity within the genes; higher α value suggests lower heterogeneity of the substitution rates among sites. For GTR+I+G analysis, the MCMC was run for 3×10^7 generations with sampling frequency of every 1,000 iterations. The convergence of MCMC runs was confirmed when the standard deviation of each split frequency below 0.01 and Potential Scale Reduction Factor equal to 1. The correlation between relative rate, α, PI sites, Pi and CI were tested using linear regression in excel.

Results

Mitochondrial Genomes

Mitochondrial genome sizes of the sister species pair, *E. decorata* (15861bps) and *E. ornata* (15863bps), are larger than that of the other sibling species pair, *E. formosa* (15700bps) and *E. yayeyamana* (15709bps) (Appendix 3). This result suggests that the variation of mitochondrial genome size in *Euphaea* damselflies has certain degree of phylogenetic component. The majority of genome size variation occurs in *A+T-rich*. The gene arrangement and composition of three newly sequenced mitochondrial genomes of *E. decorata, E. ornata* and *E. yayeyamana* are the same as that of *E. formosa* (Lin *et al*., 2010).

Pattern of Selection on nucleotide sequences

246 At the population level, the four *Euphaea* species were inferred to have the
247 same ω values of less than 1 for all examined gene sequences except for *nad5* and
248 *EF1α* (Appendix 4), indicating that most genes were under similar level of negative
249 selection across species. Nuclear *awd2* gene had an unrealistic large number of ω_T
250 (999) due to a small value of d_S (< 0.0001). For *nad5*, the ω values of the four
251 *Euphaea* species were significantly different ($\chi^2 = 23.49$, $p = 0.0001$). The least
252 populous *E. yayeyamana* has the lowest ω of *nad5* (0.0085), indicating that this
253 species had experienced the strongest negative selection in *nad5* among the four
254 *Euphaea* species; whereas *E. formosa* with a median-sized population has the highest
255 value of ω in *nad5* (0.0985), suggesting that it had experienced the weakest negative
256 selection. The ω of *EF1α* in *E. ornata* was significantly greater than 1 ($\omega = 1.25$; d_N:
257 0.0015, d_S: 0.0012), demonstrating that this gene may have experienced positive
258 selection in *E. ornata*.
259 The average value of ω of mitochondrial genes (at both population and species
260 level) were significantly lower than that of nuclear genes (at population level) ($p <$
261 0.001) (Fig. 1), suggesting that mitochondrial DNA (mtDNA) experienced stronger
262 negative selection than nuclear DNA (nuDNA). For species-level analyses of 13
263 mitochondrial protein-coding genes, the one ratio model (m0) described the variation
264 of ω values among species better than that of the four ratio model (m2) except for

atp6 (χ^2 = 57.915, p < 0.0001) (Appendix 4). All 13 mitochondrial genes were under negative selection, except for *nad3*, showing a signature of positive selection and having a ω_T of 11.6780, which was due to a small d_S value (d_N= 0.4461; d_S= 0.0382). For *atp6*, the two medium-size *Euphaea* species (*E. formosa* and *E. ornata*) have smaller ω values (0.0228 and 0.0001, respectively), whereas the most populous *E. decorata* has the largest ω (0.5742) and the least populous *E. yayeyamana* has the median value of ω (0.0422). Sliding window analyses revealed potential positive selection between site 145-151 in *awd2* gene ($d_N - d_S > 0$) (Fig. 2c). Among the eight nuclear protein-coding genes (Fig. 2, a-h), *EF1α* and *opsin* had experienced much more frequent negative selection than the other nuclear genes, which had experienced only sporadic negative selection among sites. A few nuclear codon sites had experienced stronger negative selection ($d_N - d_S < -10$) than the others (e.g., codon site 13 and 151 of *arr*, codon site 10 and 16 of *fer*, and codon site 250 of *sdhB*). The sliding windows indicated that mitochondrial genes (*cox2* and *nad5*) (Fig. 2, i & j) had also experienced negative selection at most codon sites. Overall, negative selection occurred much more often and widespread in mitochondrial than nuclear genes.

Phylogeny and sequence divergence

Species tree reconstruction revealed two pairs of sister taxa, *E. decorata* + *E.*

ornata and *E. formosa + E. yayeyamana* (Appendix. 2), a result consistent with previous studies (Huang & Lin, 2011; Lee & Lin, 2012). The shorter tree branches leading to *E. decorata* and *E. ornata* than that of *E. formosa* and *E. yayeyamana* suggest that *E. ornata* and *E. decorata* have diverged more recently than the split between *E. formosa* and *E. yayeyamana*. The pairwise sequence divergence of mitochondrial genes were approximately six times higher than that of nuclear genes (average mt: 0.0928 ± 0.0054; average nu: 0.014 ± 0.0045) (Table 1). In the comparison between each species, two sister taxa had relatively smaller pairwise mutation rates than other comparison. The sibling species, *E. decorata* and *E. ornata* had the most similar mutation rate. *Nad2* and *A+T-rich* gene had the highest mutation rate compared to other mitochondrial genes (Appendix 3).

Parsimony informative sites, consistency index

Nad2 and *A+T-rich* had the highest PI values (*nad2*= 0.129; *A+T-rich*= 0.128) among mitochondrial genes (Table 2). For nuclear genes, the intron of *opsin* and *EF1α* had the highest PI values (*opsin*= 0.057; *EF1α*= 0.056). The PI values of introns are all higher than exons in nuclear genes. The average PI values of all mitochondrial genes (0.077 ± 0.342) is two and eight fold higher than that of nuclear introns (0.039 ± 0.0158) and exons (0.009 ± 0.004), respectively, suggesting that the mitochondrial genes are phylogenetically more informative than nuclear genes at species level.

303 Mitochondrial genes had overall higher CI (0.980 ± 0.023) and on average
304 greater than that of nuclear genes (0.965 ± 0.091), but the differences was not
305 statistically significant (t= 0.561, *p*= 0.290) (Table 2). The nuclear exons appear to
306 have lower values of CI (0.973 ± 0.054). These results suggested that mitochondrial
307 genes in general had lower amount of homoplasy than that of nuclear genes.
308 *Relative substitution rates*
309 For both mitochondrial and nuclear genes, the third codon sites had the highest
310 substitution rates, followed by the first codon sites, and the second codon sites
311 contained the lowest rates, excepts for the second codon sites of *awd2* having the
312 highest rate among three codons (Fig. 3a). The relative substitution rates of nuclear
313 genes are significantly lower than that of mitochondrial genes (protein-coding and
314 non-coding) in the third codon sites ($p < 0.001$). As expected, the third codon position
315 of mitochondrial protein-coding genes had the highest substitution rates (Fig. 3a). For
316 non-coding region of mitochondrial genes, the substitution rate of *A+T-rich* is higher
317 than that of the two ribosomal genes. The lowest rate (2.0612) of the third codon site
318 of mitochondrial *nad4l* is still three times higher than the highest rate in the 3rd codon
319 site of nuclear *fer* (0.663). The highest rate of the third codon of mitochondrial *atp6*
320 (4.3913) is about 44 times higher than the lowest rate in the third codon of nuclear
321 *awd2* (0.0999). It was noticed that mitochondrial *A+T-rich* had intermediate rate

between the third codon sites of mitochondrial protein-coding and nuclear genes.

Rate heterogeneity among sites

The α (shape of gamma distribution) showed consistent differences among mitochondrial and nuclear genes. Higher values of α correspond to genes with smaller rate heterogeneity among sites (e.g. a more even distribution of rates among sites). All mitochondrial genes had higher α (approximately two fold) than that of nuclear genes ($p < 0.001$) (Fig. 3b). *A+T-rich* had the highest α (2.7769) correspond to a gene with the most even distribution of rates among sites. Among mitochondrial protein-coding genes, *nad2* has the highest α (2.0078), which is about three times higher than the lowest α in *cox2* (0.7011). Nuclear introns and exons had equivalent low α, with exon of *EF1α* the lowest (0.3485). These results showed that at species-level analyses, the mitochondrial genes have less rate heterogeneity (more even rates) among sites than that of nuclear exon and introns.

Proportion of invariant sites

Among mitochondrial genes, the highest Pi is in *sr-RNA* (0.7290), which is about five times higher than the lowest Pi in *nad2* (0.1520) (Table 2). Within nuclear genes, the highest value of Pi (*act*: 0.9724) is twice higher than the intron of *opsin* (0.4658). The average Pi of both nuclear introns (0.6884 ± 0.1459) and exons (0.9383 ± 0.0276) is higher than that of all mitochondrial genes (0.3687 ± 0.1620) ($p < 0.001$). Except

for *arr*, Pi of nuclear exons are all approaching one and larger than that of introns suggesting all sites in exons are nearly invariable.

A+T-rich had the lowest Pi (0.0920) (Fig. 3c), suggesting approximately 90% of sites are variable. The highest Pi of mitochondrial coding genes is *cox2* (0.4877), whose Pi is three times higher than the lowest Pi in *nad2* (0.1520).

Transformation rate matrices (Q marix)

There is obviously an overall higher transition rates than the transversion rates for all genes analyzed (Fig. 4, Appendix 6). Among mitochondrial genes, transition rates fall into different rates categories (AG vs. CT) in *cox1*, *cox2*, *cox3* and *nad1*, suggesting that these genes are more asymmetric in transition rates than the other mitochondrial genes with one rate category (*nad2*, *atp8*, *atp6*, *nad3-6*, *nad4l*, *cob*, *l-rRNA*, *s-rRNA* and *A+T-rich*) (Table 2, Fig. 4, Appendix 6). Whereas nuclear genes all have more symmetric transition rates (only one rate category), except for *EF1α* exon. Among transversion rates, mitochondrial *nad2*, *cox1*, *cox3*, *l-rRNA* and *s-rRNA* were more asymmetric and had two rate categories, with the remaining mitochondrial genes having one rate. In contrast, all nuclear exons and introns were inferred to have only one transversion rate category and frequently were in the same rate category as transition rates (*arr* exon, *awd2* intron, *EF1α* intron, *mlc* exon, *opsin* exon, *sdhB* exon, *sdhB* intron), indicating they had more symmetric rate matrix. The highest skew in

360 transformation rate matrix occurred in cox1 and cox3 (4 rate category), relative rate of
361 CT= 0.6394 and 0.5701 (Fig. 4, Appendix 6), which are 63 and 30 times higher than
362 the lowest rate of CG= 0.0101 and 0.0189, respectively, likely due to an excess of CT
363 transition.
364 *Correlation among parameters*
365 There was a significant negative correlation between α and Pi (r^2 = 0.6017; p <
366 0.001) (Fig. 5a), suggesting that genes have more sites grouped into invariant sites
367 (higher Pi), the remaining sites tend to show more rate heterogeneity (lower α). Alpha
368 (α), the shape parameter of gamma distribution, showed a significant positive
369 correlation with PI sites (r^2 = 0.6112; p < 0.001), suggesting that genes with less rate
370 heterogeneity (higher α) such as mitochondrial *A+T-rich* and *nad2* has more
371 informative sites than many mitochondrial and nuclear genes (Fig. 5b). Although the
372 majority of mitochondrial genes and nuclear introns have a base composition of
373 approximately 70% A+T. The relative substitution rate had a significant positive
374 correlation with base composition (A+T %) (r^2 = 0.4944; p < 0.001) (Fig. 5c). PI, the
375 parsimony informative sites, was positively correlated with relative substitution rate
376 (r^2 = 0.6721; p < 0.001) (Fig. 5d), suggesting that the genes with higher relative
377 substitution rate showed also have higher informative sites. The correlation between
378 CI, the consistency index, has both not significant with relative substitution rate (r^2 =

379 0.0001; $p = 0.8614$) and alpha (α) ($r^2 = 0.0066$; $p = 0.6636$) because all CI values

380 approach 1.

381

382 **Discussion**

383 Our results indicated that the majority of mitochondrial and nuclear genes

384 experienced frequent negative selection, with exception of potential positive selection

385 in *EF1α* at species level and *nad3* at population level. The main selective force

386 driving the sequence variation in mitochondrial and nuclear genes was negative

387 selection rather than positive selection (genetic draft) as proposed by earlier studies

388 (Bazin *et al*., 2006, Mulligan *et al*., 2006). The proteins transcribed by mitochondrial

389 genes are important in energy metabolism and related to the electron transport chains,

390 therefore the mutation in mitochondrial genes was expected to be under strong

391 negative selection. Most mutations in mitochondrial and nuclear coding genes were

392 likely deleterious and removed by negative selection from evolving lineages

393 (Eyre-Walker & Keightley, 2007; Martincorena & Luscombe, 2012). We found

394 evidences to suggest that mitochondrial genes had experienced stronger negative

395 selection than nuclear genes, despite population sizes of mitochondrial genes are one

396 fourth of that of nuclear genes. However, the strength of negative selection did not

397 correlate with the population sizes, suggesting that population sizes of *Euphaea*

398 species did not predict the strength of selection on sequence variation. Therefore, the

399 comparative analysis of pattern of selection in mitochondrial and nuclear genes of
400 *Euphaea* species does not support the hypothesis that genetic draft by positive
401 selection is more frequent in species with larger population sizes and in genes with
402 smaller effective population sizes such as mitochondrial genes.

403 Although the barcoding region of COI gene is a widely used genetic marker in
404 species identification (over 464 thousand reference sequences of animals in the
405 Barcode of Life Data System, http://www.boldsystems.org), the successful
406 identification rate for closely related species were low (Hebert *et al*., 2003; Elias *et al*.,
407 2007; Rockney *et al*., 2015). Studies showed that the variation of this region was
408 better used for identifying distantly related species (Vences *et al*., 2005d) rather than
409 distinguishing sister taxa (Meier *et al*., 2006). According to our data, the barcoding
410 region of COI gene also had a medium substitution rate, α and Pi ratio among all
411 mitochondrial genes.

412 Previous study showed that the fixation happened in species can cause unique
413 haplotype for species delimitation (Roe & Sperling, 2007b). Another study also
414 showed that the mtDNA and ribosomal DNA had higher haplotype fixation than
415 nuDNA (Dupuis *et al.*, 2012), suggesting these two gene groups will be more efficient
416 in species delimitation. Among mitochondrial and nuclear genes, mitochondrial
417 protein-coding gene, *nad2* and non-coding *A+T-rich* exhibit higher phylogenetic

418 utility than the others. A recent study of choosing genetic markers showed that some
419 mtDNA (*nad1*, *nad2*, *nad4*, *nad5* and *l-rRNA*) recovered more monophyletic insect
420 taxa (McDonagh *et al*., 2016). A previous study suggested that mitochondrial control
421 region (*A+T-rich*) is too conserve to be a good marker for constructing deep
422 intraspecies phylogeny, but it can delimitate close sister taxa (Zhang & Hewitt, 1997).
423 According to our result, *A+T-rich* exhibits the best quality for phylogenetic analyses,
424 including the highest α, relative rate at species level, which suggesting *A+T-rich*
425 carries more information than other mtDNA. Also, the high relative substitution rate
426 of *A+T-rich* suggests that can solve relationship of recent diverged species.
427 Previous studies mainly focus on selecting genetic markers suitable for
428 constructing deep phylogenetic splits of insect taxa (Lin & Danforth, 2004; Danforth
429 *et al*., 2005; McDonagh *et al*., 2016). Our study provided a comprehensive ranking of
430 mitochondrial genes for phylogenetic and population genetics analyses of odonates at
431 species level. This information can provide a useful guide for the selection of
432 mitochondrial genes that best fit the level of divergence in insects of interest.

433 **Acknowledgement**

434 We thank Ming-Yu Chen for sequencing mitochondrial genomes; Jo-Fan Wang
435 for helps of data analyses. Jen-Pen Huang for provides specimens. This project was
436 supported by the National Science Council of Taiwan (NSC

100-2311-B-029-004-MY3).

References

Ballard, J.W.O. and Whitlock, M.C. (2004) The incomplete natural history of mitochondria. *Mol Ecol* **13**: 729–744.

Bazin, E., Glémin, S., and Galtier, N. (2006) Population size does not influence mitochondrial genetic diversity in animals. *Science* **312**: 570–572.

Danforth, B.N., Sipes, S., Fang, J. and Brady, S.G. (2006) The history of early bee diversification based on five genes plus morphology. *P Natl Acad Sci* **103**: 15118–15123.

Drummond, A.J., Suchard, M.A., Xie, D. and Rambaut, A. (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29**: 1969–1973.

Dupuis, J.R., Roe, A.D. and Sperling, F.A.H. (2012) Multi-locus species delimitation in closely related animals and fungi: one marker is not enough. *Mol Ecol* **21**: 4422–4436.

Elias, M., Hill, R.I., Willmott, K.R., Dasmahapatra, K.K., Brower, A.V.Z., Mallet, J.

459 and Jiggins, C.D. (2007) Limited performance of DNA barcoding in a diverse
460 community of tropical butterflies. *Proc R Soc B* **274**: 2881–2889.
461
462 Eyre-Walker, A. and Keightley, P.D. (2007) The distribution of fitness effects of new
463 mutations. *Nature* **8**: 610–618.
464
465 Frankham, R. (1996) Relationship of genetic variation to population size in wildlife.
466 *Conserv Biol* **10**: 1500–1508.
467
468 Galtier, N., Nabholz, B., Glémin, S., and Hurst, G.D.O. (2009) Mitochondrial DNA as
469 a marker of molecular diversity: a reappraisal. *Mol Ecol* **18**: 4541–4550.
470
471 Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D. and Bairoch, A.
472 (2003) ExPASy: the proteomics server for in-depth protein knowledge and
473 analysis. *Nucleic Acids Res* **31**: 3784–3788.
474
475 Halliburton, R. (2004) *Introduction to population genetics*. Pearson Education Inc,
476 Upper Saddle River, New Jersey.
477

Hauser, L., Adcock, G.J., Smith, P.J., Ramírez, J.H.B. and Carvalho, G.R. (2002) Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*). *P Natl Acad Sci* **99**: 11742–11747.

Hebert, P.D.N., Ratnasingham, S. and deWaard, J.R. (2003) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *T Roy Soc* **270**: S96–S99.

Heller, R., Okello, J.B.A. and Siegismund, H. (2010) Can small wildlife conservancies maintain genetically stable populations of large mammals? Evidence for increased genetic drift in geographically restricted populations of Cape buffalo in East Africa. *Mol Ecol* **19**: 1324–1334.

Huang, J.P. and Lin, C.P. (2011) Lineage-specific late pleistocene expansion of an endemic subtropical gossamer-wing damselfly, *Euphaea formosa*, in Taiwan. *BMC Evol Biol* doi: 10.1186/1471-2148-11-94.

Huelsenbeck, J.P. and Ronquist, F.R. (2001) MrBayes: Bayesian inference of

497 phylogenetic trees. *Bioinformative* **17**: 754–755.

498

499 Lee, Y.H. and Lin, C.P. (2012) Pleistocene speciation with and without gene flow in
500 *Euphaea* damselflies of subtropical and tropical East Asian islands. *Mol Ecol*
501 **21**: 3739–3756.

502

503 Librado, P. and Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of
504 DNA polymorphism data. *Bioinformative* **25**: 1451–1452.

505

506 Lin, C.P. and Danforth, B.N. (2004) How do insect nuclear and mitochondrial genes
507 substitution patterns differ? Insights from Bayesian analysis of combined
508 datasets. *Mol Phylogenet Evol* **30**: 686–702.

509

510 Lin, C.P., Chen, M.Y. and Huang, J.P. (2010) The complete mitochondrial genome
511 and phylogenomics of a damselfly, *Euphaea formosa* support a basal Odonata
512 within the Pterygota. *Gene* **468**: 20–29.

513

514 Luikart, G., Ryman, N., Tallmon, D.A., Schwartz, M.K. and Allendorf, F.W. (2010)
515 Estimation of census and effective population sizes: the increasing usefulness

516 of DNA-based approaches. *Conserv Genet* **11**: 355–373.

517

518 Martincorena, I. and Luscombe, N.M. (2012) Non-random mutation:the evolution of

519 targeted hypermutation and hypomutation. *Bioessays* **35**: 123–130.

520

521 McDonagh, L.M., West, H., Hrrison, J.W. and Stevens, J.R. (2016) Which

522 mitochondrial gene (if any) is best for insect phylogenetics? *Insect Syst Evol*

523 doi:10.1163/1876312X-47032142

524

525 Meiklejohn, C.D., Montooth, K.L. and Rand, D.M. (2007) Positive and negative

526 selection on the mitochondrial genome. *Trends Genet* **23**: 259–263.

527

528 Meier, R., Shiyang, K., Vaidya, G. and NG, P.K.L. (2006) DNA Barcoding and

529 Taxonomy in Diptera: A Tale of High Intraspecific Variability and Low

530 Identification Success. *Syst Biol* **55**: 715–728.

531

532 Miyata, T. and Yasunaga, T. (1980) Molecular evolution of mRNA: a method for

533 estimating evolutionary rates of synonymous and amino acid substitutions from

534 homologous nucleotide sequences and its application. *J Mol Evol* **16**: 23–36.

Miller, C.R. and Waits, L.P. (2003) The history of effective population size and genetic diversity in the Yellowstone grizzly (*Ursus arctos*): implcations for conservation. *Proc Natl Acad Sci USA* **100**: 4334–4339.

Moriyama, E.N. and Gojobori, T. (1992) Rates of synonymous substitution and base composition of nuclear genes in Drosophila. *Genetics* **130**: 855–864.

Mulligan, C.J., Kitchen, A. and Miyamoto, M.M. (2006) Comment on "population size does not influence mitochondrial genetic diversity in animals". *Science* **314**: 1390.

Nielsen, R. and Slatkin, M. (2013) *An introduction to population genetics*. Sinauer Associates Inc, Sunderland, Massachusetts.

Papadopoulou, A., Anastasiou, I. and Vogler, A.P. (2010) Revisiting the insect mitochondrial molecular clock: the mid-aegean trench calibration. *Mol Biol Evol* **27**: 1659–1672.

554 Piganeau, G. and Eyre-Walker, A. (2009) Evidence for variation in the effective
555 population size of animal mitochondrial DNA. *PLoS One* **4**: e4396.
556
557 Rambaut, A., Suchard, M., Xie, D. and Drummond, A. (2014) Tracer v1.6. Available
558 from http://beast.bio.ed. ac.uk/Tracer.
559
560 Rockney, H.J., Ofori-Boateng, C., Porcino, N. and Leaché, A.D. (2015) A comparison
561 of DNA barcoding markers in West African frogs. *Afr J Herpetol* **64**: 135–147.
562
563 Roe, A.D. and Sperling, F.A.H. (2007b) Blackwell publishing ltd population structure
564 and species boundary delimitation of cryptic *Dioryctria* moths: an integrative
565 approach. *Mol Ecol* **16**: 3617–3633.
566
567 Roman, J. and Palumbi, S.R. (2003) Whales before whaling in the North Atlantic.
568 *Science* **301**: 508–510.
569
570 Roux, J., Privman, E., Moretti, S., Daub, J.T., Robinson-Rechavi, M. and Keller, L.
571 (2014) Patterns of Positive Selection in Seven Ant Genomes. *Mol Biol Evol* **31**:
572 1661–1685.

573

574 Simon, C., Frati F., Beckenbach, A., Crespi, B., Liu, H. and Flook, P. (1994)
575 Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences
576 and a compilation of conserved polymerase chain reaction primers. *Ann*
577 *Entomol Soc Am* **87**: 651–701.

578

579 Simon, C., Bucklry, T.R., Frati, F. Stewart, J.B. and Beckenbach, A.T. (2006)
580 Incorporating molecular evolution into phylogenetic analysis, and a new
581 compilation of conserved polymerase chain reaction primers for animal
582 mitochondrial DNA. *Annu Rev Ecol Evol Syst* **37**: 545–579.

583

584 Tajima, F. and Nei, M. (1984). Estimation of evolutionary distance between
585 nucleotide sequences. *Mol Biol Evol* **1**: 269–285.

586

587 Tamura, K., Peterson, D., Peterson, N., Stecher G., Nei, M. and Kumar, S. (2011)
588 MEGA5: Molecular evolutionary genetics analysis using maximum likelihood,
589 evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**:
590 2731–2739.

591

592 Tamura, K., Stecher G., Peterson, D., Filipski, A. and Kumar, S. (2013) MEGA6:
593 Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30**: 2725–
594 2729.
595
596 Vences, M., Thomas, M., Bonett, R.M. and Vieites, D.R. (2005) Deciphering
597 amphibian diversity through DNA barcoding: chances and challenges. *Philos T*
598 *Roy Soc B* **360**: 1859–1868.
599
600 Wright, S. (1931) Evolution in Mendelian populations. Genetics **16**: 97–159.
601
602 Yamauchi, M.M., Miya, M.U. and Nishida, M. (2004) Use of a PCR-based approach
603 for sequencing whole mitochondrial genomes of insects: two example
604 (cockroach and dragonfly) based on the method developed for decapod
605 crustaceans. *Insect Mol Biol* **13**: 435–442.
606
607 Yang, Z. and Bielawski, J.P. (2000) Statistical methods for detecting molecular
608 adaptation. *Trends Ecol Evol* **15**: 496–503.
609
610 Yang, Z. (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol*

611 *Evol* **24**: 1586–1591.

612

613 Zhang, D. and Hewitt, G.M. (1997) Insect mitochondrial control region: A review of

614 its structure, evolution and usefulness in evolutionary studies. *Biochem Syst*

615 *Ecol* **25**: 99–120.

Table 1. Mutation rates of mitochondrial and nuclear genes between *Euphaea* species calculated using Tajima-Nei model in MEGA. The mutation rate of 1.77 × 10^{-8} and 1.2 × 10^{-8} mutation /site/year of *cox1* and *arr* were used for calibration of sequence divergence in mitochondrial and nuclear genes, respectively.

Gene	*E. y.* vs *E. d.*	*E. y.* vs *E. o.*	*E. f.* vs. *E. d.*	*E. f.* vs *E. o.*	*E. f.* vs *E. y.*	*E. d.* vs *E. o.*	Average
Nuclear							
act	0.0087 ± 0.0028	0.0087 ± 0.0028	0.0084 ± 0.0014	0.0084 ± 0.0014	0.0063 ± 0.0028	0.0012 ± 0.0014	0.028 ± 0.007
arr	0.0254 ± 0.0075	0.0227 ± 0.0047	0.0393 ± 0.0075	0.0345 ± 0.0073	0.0218 ± 0.0065	0.0244 ± 0.0084	0.0069 ± 0.0021
awd2	0.0033 ± 0.0038	0.0055 ± 0.005	0.0055 ± 0.005	0.0059 ± 0.0052	0.0055 ± 0.005	0.0055 ± 0.005	0.0099 ± 0.0022
EF1a	0.0137 ± 0.0017	0.0112 ± 0.0025	0.0138 ± 0.0016	0.0114 ± 0.0024	0.0044 ± 0.0022	0.0048 ± 0.0029	0.0197 ± 0.0054
fer	0.0176 ± 0.0048	0.0223 ± 0.0063	0.0224 ± 0.0047	0.0233 ± 0.0059	0.0164 ± 0.0035	0.0162 ± 0.0069	0.0121 ± 0.0061
mlc	0.0189 ± 0.0078	0.0168 ± 0.0064	0.0125 ± 0.0063	0.0104 ± 0.0036	0.0062 ± 0.0063	0.0076 ± 0.0061	0.0127 ± 0.0028
opsin	0.0177 ± 0.0042	0.0177 ± 0.0034	0.0153 ± 0.0034	0.0159 ± 0.0014	0.0033 ± 0.0021	0.006 ± 0.0023	0.0174 ± 0.006
sdhB	0.0185 ± 0.0063	0.0164 ± 0.0052	0.0224 ± 0.007	0.0204 ± 0.0059	0.0211 ± 0.0063	0.0059 ± 0.0051	0.0052 ± 0.0048
Average	0.0155 ± 0.0049	0.0152 ± 0.0045	0.0175 ± 0.0046	0.0163 ± 0.0041	0.0106 ± 0.0043	0.0089 ± 0.0047	0.014 ± 0.0045
Mitochondrial							
cox2	0.1042 ± 0.0058	0.0978 ± 0.0062	0.1093 ± 0.0038	0.1086 ± 0.0045	0.0689 ± 0.0137	0.0136 ± 0.0037	0.0837 ± 0.0063
nad5	0.1395 ± 0.0036	0.1357 ± 0.003	0.1302 ± 0.0028	0.1304 ± 0.0028	0.0593 ± 0.0125	0.0158 ± 0.0022	0.1018 ± 0.0045
Average	0.1219 ± 0.0047	0.1167 ± 0.0046	0.1198 ± 0.0033	0.1195 ± 0.0037	0.0641 ± 0.0131	0.0147 ± 0.0029	0.0928 ± 0.0054

620 **Table 2.** Summary of parameters of phylogenetic analyses of mitochondrial and nuclear genes. PI sites:

621 parsimony informative sites; CI: consistency index, Pi: proportion of invariant sites, α: shape of gamma

622 distribution. Substitution rate categories refer to Fig. 4 and Appendix 6.

Gene	Length	A+T%	PI sites	CI[a]	Pi	α	Substitution rate model		
							TS	TV	total
Mt coding									
nad2	990	74.1	0.129	0.984	0.1520	2.0078	a	b, c	3
cox1	1548	64.3	0.089	0.972	0.3411	1.1163	a, b	c, d	4
cox2	688	67.8	0.044	1	0.4877	0.7011	a, b	c	3
atp8	159	77.2	0.006	1	0.3157	1.2044	a	b	2
atp6	675	69.9	0.070	1	0.2695	1.3661	a	b	2
cox3	787	64.6	0.088	0.986	0.3081	1.1830	a, b	c, d	4
nad3	354	70.5	0.118	0.953	0.2721	1.3405	a	b	2
nad5	1723	70.3	0.076	0.978	0.3604	0.7979	a	b	2
nad4	1344	71.6	0.081	0.956	0.3897	1.0041	a	b	2
nad4l	294	73.4	0.069	0.952	0.4180	0.8899	a	b	2
nad6	498	73.0	0.103	1	0.2657	1.3477	a	b	2
cob	1134	68.5	0.045	0.927	0.4712	0.7676	a	b	2
nad1	951	69.1	0.073	0.958	0.3831	1.0008	a, b	c	3
Mean ± SD		70.3 ± 3.7	0.076 ± 0.0324	0.9743 ± 0.0236	0.3411 ± 0.0921	1.1329 ± 0.3467			
Mt non-coding									
l-rRNA	1291	68.9	0.053	1	0.6433	0.7835	a	b, c	3
s-rRNA	781	73.9	0.050	1	0.7290	0.7498	a	b, c	3
A+T-rich	1109	81.6	0.128	0.993	0.0920	2.7769	a	b	2
Mean ± SD		74.8 ± 6.4	0.077 ± 0.0442	0.9977 ± 0.004	0.4881 ± 0.3457	1.4367 ± 1.1607			
Nu exon									
act	423	44.0	0.007	1	0.9724	0.5120	a	b	2
arr	303	45.9	0.007	1	0.8836	0.6937	a	a	1
awd2	153	54.9	0.007	1	0.9566	0.7341	a	b	2
EF1α	708	48.1	0.007	1	0.9543	0.3485	a, b	c	3
fer	333	52.9	0.018	0.667	0.9219	0.7074	a	b	2
mlc	162	46.9	0.006	1	0.9513	0.6243	a	a	1
opsin	309	46.3	0.01	1	0.9405	0.6161	a	a	1
sdhB	200	55.5	0.01	1	0.9261	0.6304	a	a	1
Mean ± SD		49.3 ± 4.4	0.009 ± 0.0039	0.9584 ± 0.1177	0.9383 ± 0.0276	0.6083 ± 0.1256			
Nu intron									

arr	1039	60.1	0.038	0.952	0.6967	0.7525	a	b	2
awd2	211	76.8	0.043	1	0.7687	0.6882	a	a	1
EF1α	252	68.5	0.056	1	0.6061	0.7656	a	a	1
fer	879	64.3	0.043	1	0.7612	0.7358	a	b	2
mlc	731	68.9	0.016	0.857	0.6030	0.6811	a	b	2
opsin	1075	69.4	0.057	1	0.4658	0.8496	a	b	2
sdhB	200	71.5	0.02	1	0.9170	0.5991	a	a	1
Mean ± SD		68.5 ± 5.3	0.039 ± 0.016	0.9728 ± 0.0540	0.6884 ± 0.1459	0.7246 ± 0.0788			

[a] Excluding un-informative sites

626 **Figure 1.** The ω (nonsynonymous/ synonymous substitution ratio, d_N/d_S) of nuclear

627 and mitochondrial genes on population and species level. The ω of nuclear genes

628 (0.102 ± 0.0451) is significantly different with that of mitochondrial genes at

629 population (0.003 ± 0.0336) and species (0.029 ± 0.0812) level (Tukey test).

630 **Figure 2.** Sliding window analyses of $d_N - d_S$ of nuclear (a-h) and mitochondrial (i-j)

631 genes. The arrows indicate codon sites with $d_N - d_S$ smaller than –10.

632 **Figure 3.** (a) Relative rate among codon sites of mitochondrial protein-coding,

633 non-coding and nuclear genes estimated in GTR+SSR model. (b) The shape of

634 gamma distribution (α) of mitochondrial protein-coding, non-coding and nuclear

635 genes estimated in GTR+I+G model. (c) The proportion of invariant sites (Pi) of

636 mitochondrial protein-coding, non-coding and nuclear genes estimated in GTR+I+G

637 model.

638 **Figure 4.** Examples of the transformation rate matrix in selected mitochondrial and

639 nuclear genes (mean ± 95% credibility interval, CI). The alphabet above the bars

640 represents the rate categories estimated in model selection of MEGA.

641 **Figure 5.** (a) Relationship between α (the shape of gamma distribution) and Pi

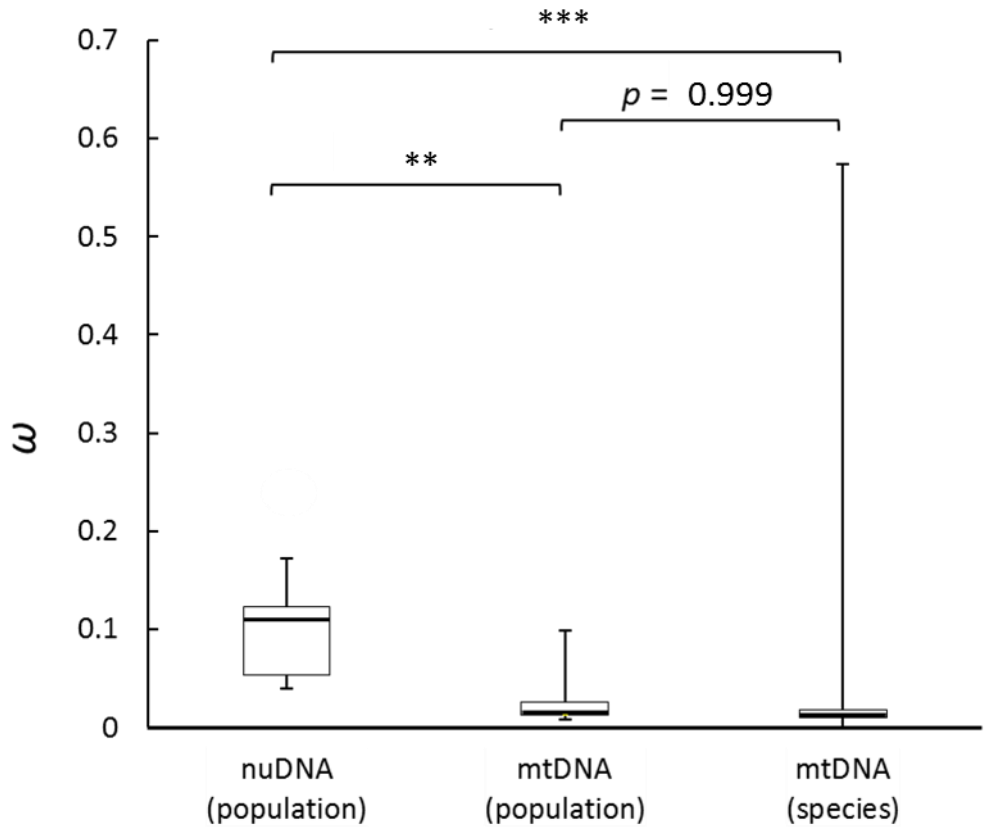
642 (estimated proportion of invariant sites). (b) Relationship between α and PI

643 (parsimony informative sites). (c) Relationship between relative substitution rate and

644 A+T bias. (d) Relationship between PI and relative rate. All regressions are significant

645 ($P < 0.001$).

646 Fig. 1

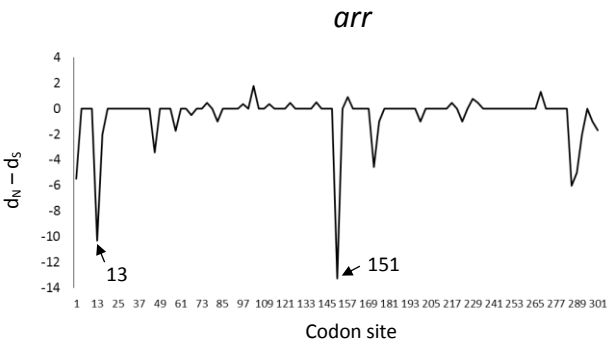

***
p = 0.999
**
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
ω
nuDNA
(population)
mtDNA
(population)
mtDNA
(species)

647

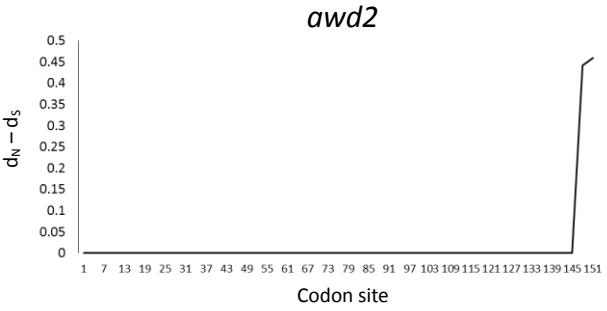
Fig. 2

(a)

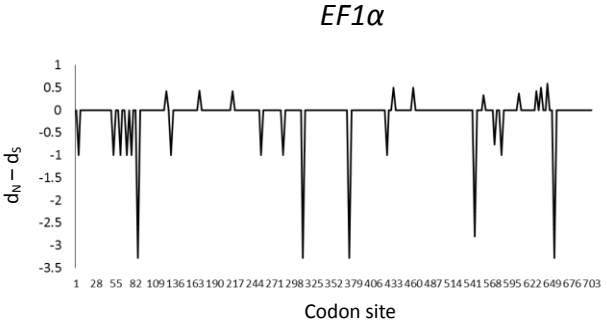
(b)



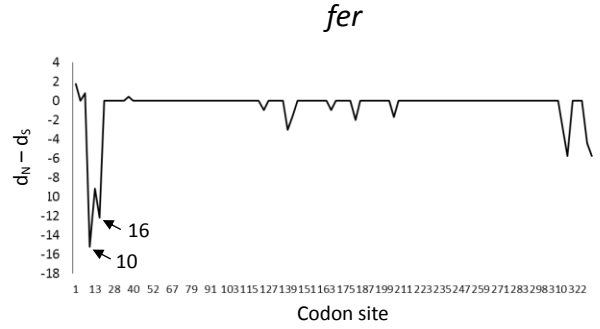
(c)



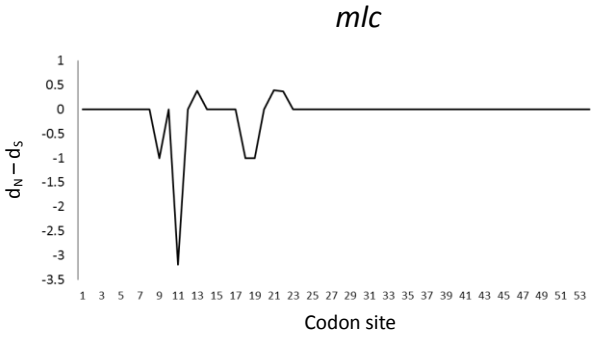
(d)



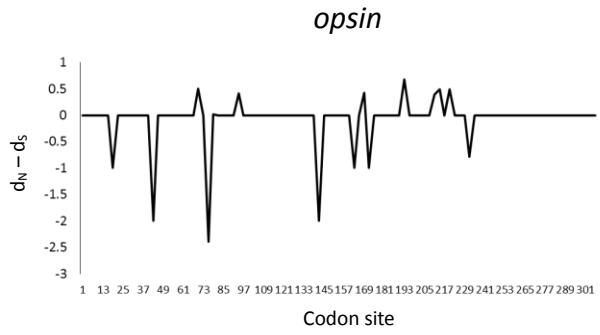
(e)



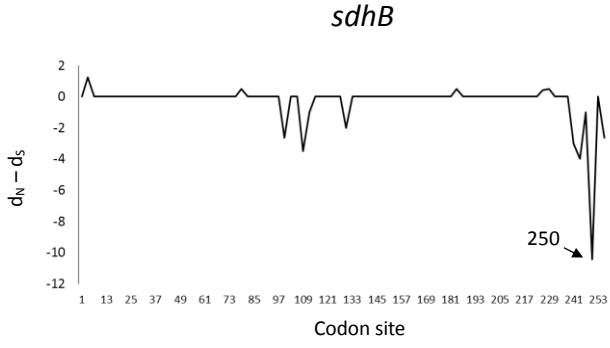
(f)



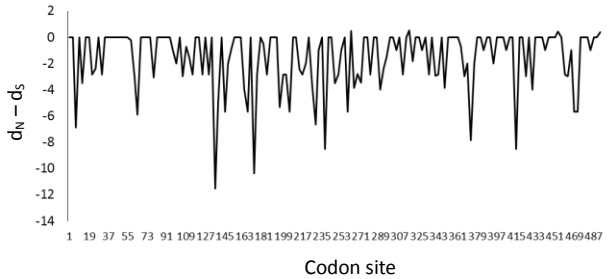
(g)



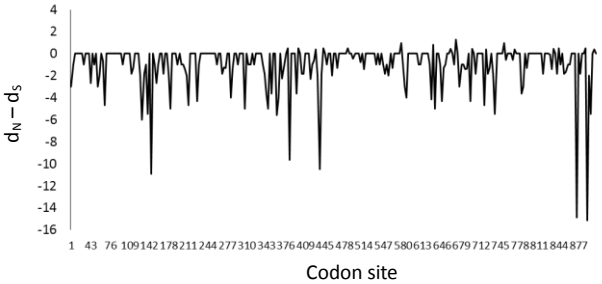
(h)



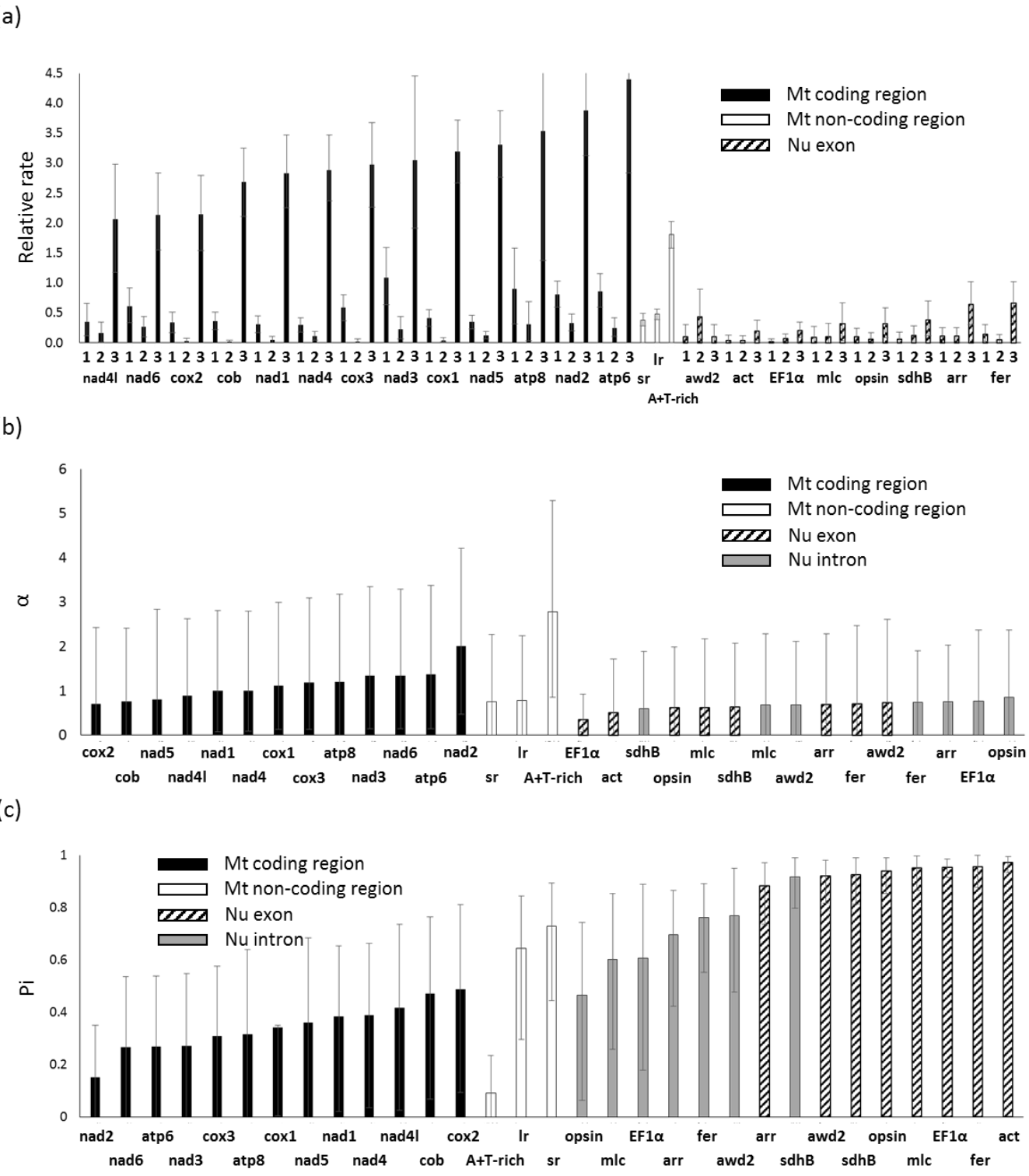
(i)



(j)

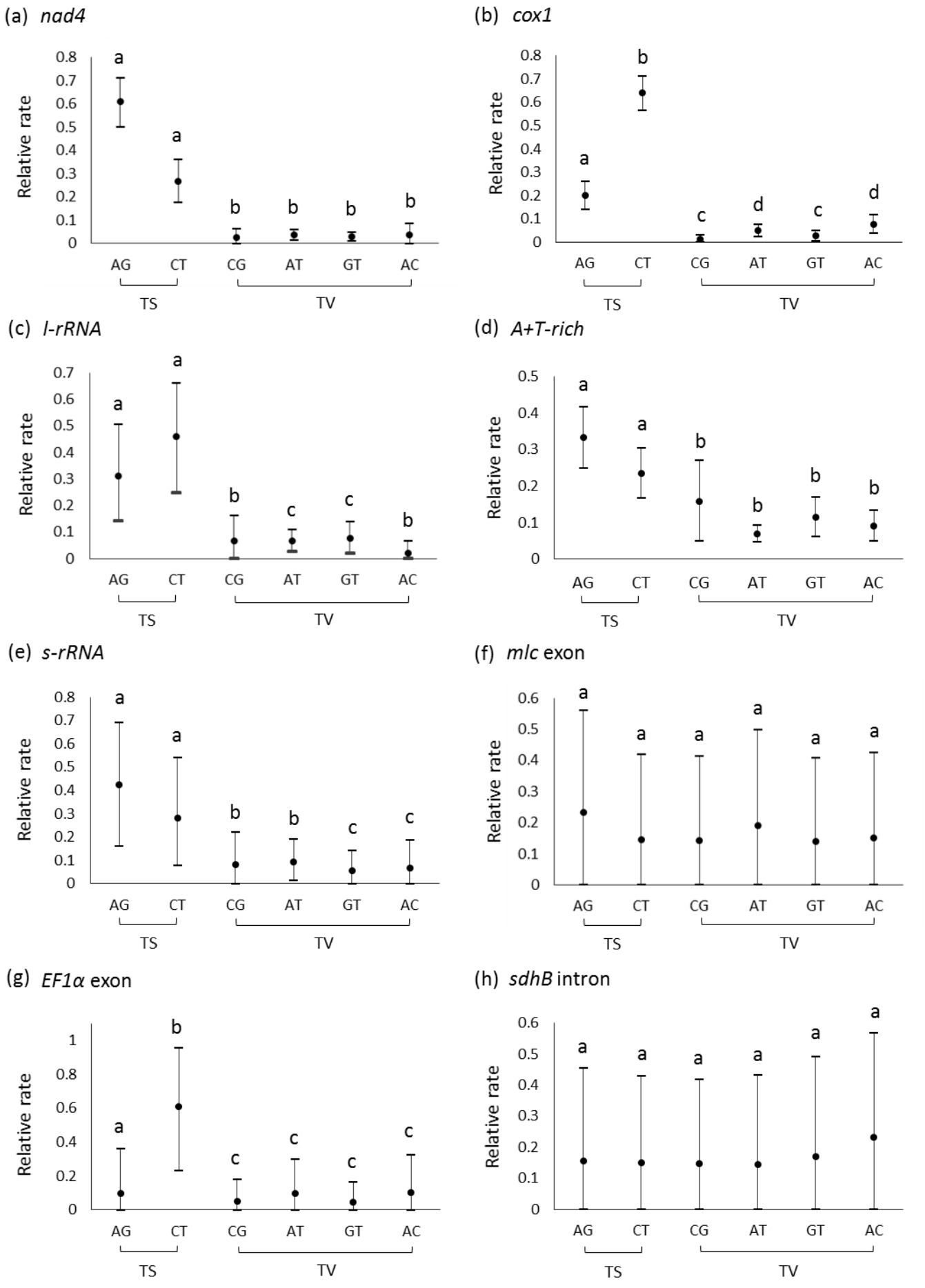


659 Fig. 3


(a)
Relative rate
Mt coding region
Mt non-coding region
Nu exon
nad4l nad6 cox2 cob nad1 nad4 cox3 nad3 cox1 nad5 atp8 nad2 atp6 sr lr A+T-rich awd2 act EF1α mlc opsin sdhB arr fer
(b)
α
Mt coding region
Mt non-coding region
Nu exon
Nu intron
cox2 cob nad5 nad4l nad1 nad4 cox1 cox3 atp8 nad3 nad6 atp6 nad2 sr lr A+T-rich EF1α act sdhB opsin mlc sdhB mlc awd2 arr fer awd2 fer arr EF1α opsin
(c)
Pi
Mt coding region
Mt non-coding region
Nu exon
Nu intron
nad2 nad6 atp6 nad3 cox3 atp8 cox1 nad5 nad1 nad4 nad4l cob cox2 A+T-rich lr sr opsin mlc EF1α arr fer awd2 arr sdhB awd2 sdhB opsin mlc EF1α fer act

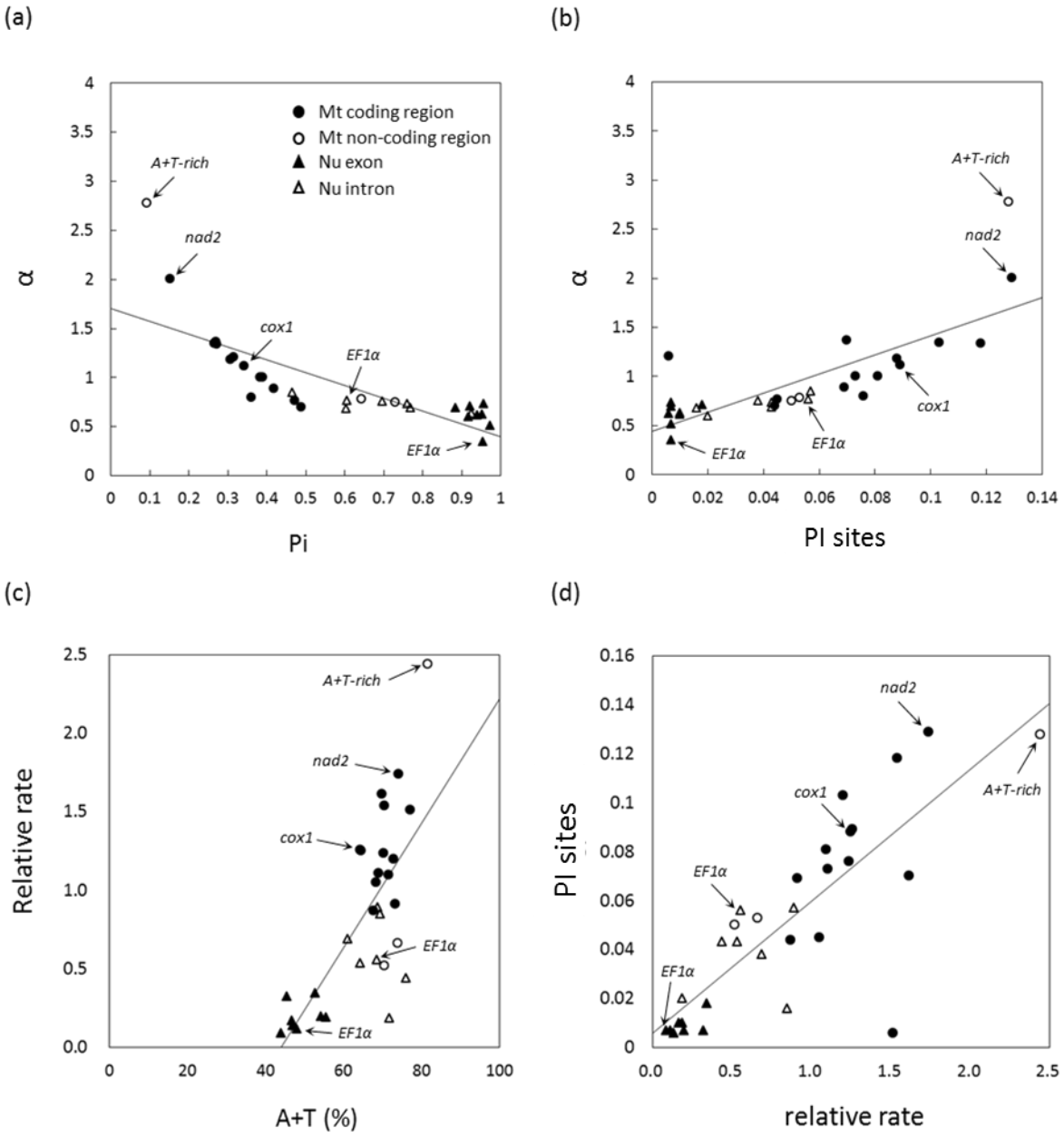
660

661 Fig. 4


(a) nad4
(b) cox1
(c) l-rRNA
(d) A+T-rich
(e) s-rRNA
(f) mlc exon
(g) EF1α exon
(h) sdhB intron
Relative rate
AG CT CG AT GT AC
TS TV

662

663 Fig. 5



664

665

666

667 **Appendix 1.** Information of specimens and GenBank accession numers of two

668 mitochondrial and eight nuclear genes used in this study.

Gene	Species	Accession numbers (Lee & Lin 2012; Haung & Lin 2011)	Haplotype numbers in Haung & Lin 2011	Non-submitted species number
cox2	*E. yayeyamana*	JF918938.1		
		JF918959.1		
		JF918956.1		
		JF918955.1		
		JF918953.1		
		JF918951.1		
		JF918941.1		
	E. formosa	EU603537.1	H46	
		EU603535.1	H01	
		EU603538.1	H20	
		EU603527.1	H17	
		EU603539.1	H39	
		EU603540.1	H24	
		EU603541.1	H14	
		EU603563.1	H31	
		EU603562.1	H33	
		EU603561.1	H36	
		EU603559.1	H49	
		EU603558.1	H45	
		EU603556.1	H26	
	E. ornata	JN793716.1		
		JN793717.1		
		JN793718.1		
		JN793719.1		
		JN793721.1		
		JN793724.1		
		JN793728.1		
		JN793729.1		
		JN793731.1		
	E. decorata	JN793699.1		

		JN793700.1	
		JN793704.1	
		JN793710.1	
		JN793713.1	
		JN793715.1	
			EfAa12wi
nad5	*E. yayeyamana*	JQ974328.1	
		JQ974330.1	
		JQ974332.1	
		JQ974334.1	
		JQ974335.1	
		JQ974337.1	
		JQ974339.1	
		JQ974340.1	
		JQ974341.1	
		JQ974343.1	
	E. formosa	JQ974312.1	
		JQ974313.1	
		JQ974314.1	
		JQ974315.1	
		JQ974316.1	
		JQ974318.1	
		JQ974319.1	
		JQ974320.1	
		JQ974321.1	
		JQ974326.1	
		JQ974327.1	
	E. ornata	JQ974360.1	
		JQ974361.1	
		JQ974362.1	
		JQ974363.1	
		JQ974364.1	
		JQ974365.1	
		JQ974367.1	
		JQ974368.1	
		JQ974370.1	
		JQ974372.1	
		JQ974373.1	

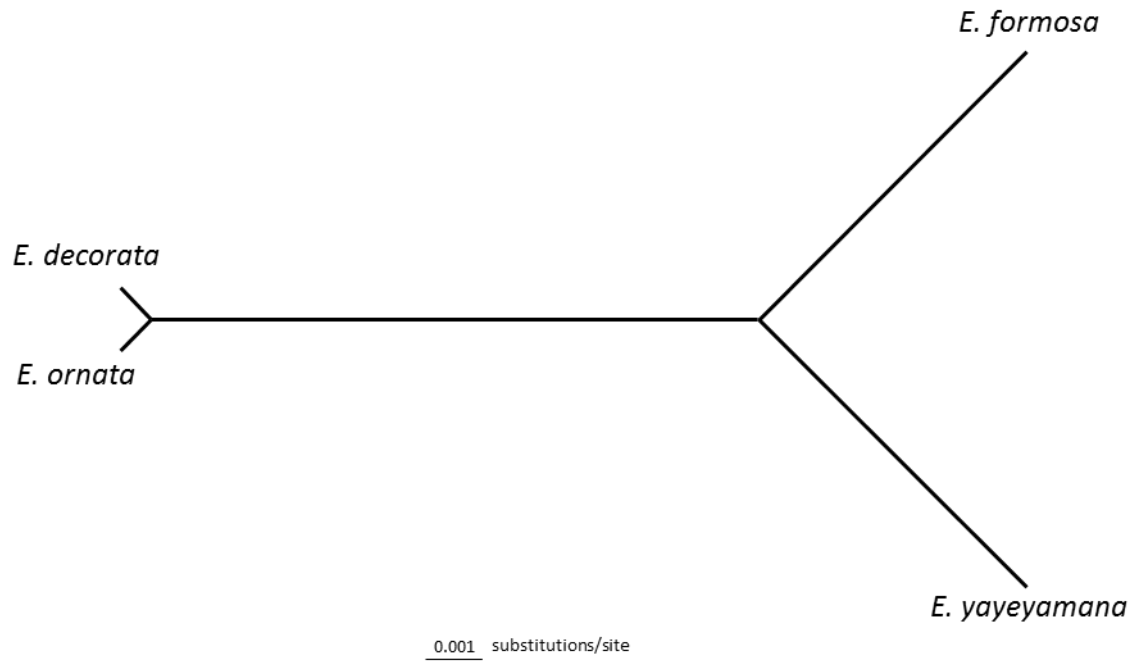
		JQ974374.1	
	E. decorata	JQ974344.1	
		JQ974350.1	
		JQ974353.1	
		JQ974354.1	
		JQ974356.1	
		JQ974357.1	
		JQ974359.1	
			EfAa1
			EfAa2
			EfBa1
			EfBa5
			EfBa12
			EfDa4
			EfTa3
			TW1
act	*E. yayeyamana*	JN389860.1	
		JN389857.1	
	E. formosa	JN389838.1	
		JN389839.1	
		JN389843.1	
		JN389854.1	
	E. ornata	JN389814.1	
		JN389818.1	
	E. decorata	JN389796.1	
		JN389799.1	
			EfAa1
			EfBa2
			EfBa5
arr	*E. yayeyamana*	JN389890.1	
		JN389891.1	
		JN389892.1	
		JN389894.1	
		JN389896.1	
		JN389898.1	
		JN389899.1	
	E. formosa	JN389875.1	
		JN389876.1	

		JN389877.1
		JN389883.1
	E. ornata	JN389930.1
		JN389932.1
		JN389933.1
		JN389937.1
		JN389938.1
		JN389939.1
		JN389941.1
		JN389942.1
		JN389943.1
	E. decorata	JN389909.1
		JN389910.1
		JN389911.1
		JN389912.1
		JN389913.1
		JN389914.1
		JN389915.1
		JN389916.1
		JN389917.1
		JN389918.1
		JN389923.1
awd2	*E. yayeyamana*	JN246946.1
		JN246952.1
		JN246960.1
	E. formosa	JN246927.1
		JN246928.1
	E. ornata	JN246984.1
		JN246987.1
		JN246997.1
	E. decorata	JN246963.1
		JN246967.1
EF1α	*E. yayeyamana*	JN389968.1
		JN389969.1
		JN389971.1
		JN389974.1
		JN389975.1
	E. formosa	JN389950.1

		JN389951.1	
		JN389952.1	
	E. ornata	JN390008.1	
		JN390014.1	
		JN390022.1	
	E. decorata	JN389988.1	
		JN389993.1	
		JN390003.1	
			EfAa9
fer	*E. yayeyamana*	JN390052.1	
		JN390055.1	
		JN390058.1	
		JN390076.1	
	E. formosa	JN390028.1	
		JN390032.1	
		JN390038.1	
		JN390046.1	
		JN390047.1	
		JN390051.1	
	E. ornata	JN390097.1	
		JN390099.1	
		JN390105.1	
		JN390106.1	
		JN390107.1	
	E. decorata	JN390077.1	
		JN390079.1	
		JN390080.1	
		JN390082.1	
		JN390090.1	
		JN390091.1	
		JN390093.1	
		JN390094.1	
mlc	*E. yayeyamana*	JN390133.1	
	E. formosa	JN390111.1	
		JN390119.1	
		JN390125.1	
	E. ornata	JN390172.1	
		JN390179.1	

		JN390191.1
	E. decorata	JN390153.1
		JN390157.1
		JN390168.1
opsin	*E. yayeyamana*	JN390214.1
	E. formosa	JN390196.1
		JN390197.1
		JN390200.1
		JN390202.1
		JN390204.1
		JN390206.1
	E. ornata	JN390253.1
		JN390254.1
		JN390255.1
		JN390256.1
		JN390261.1
		JN390263.1
	E. decorata	JN390230.1
		JN390231.1
		JN390232.1
		JN390234.1
		JN390240.1
		JN390249.1
sdhB	*E. yayeyamana*	JN390295.1
		JN390296.1
		JN390297.1
	E. formosa	JN390274.1
		JN390275.1
		JN390276.1
		JN390277.1
		JN390278.1
		JN390279.1
		JN390281.1
		JN390283.1
		JN390291.1
	E. ornata	JN390326.1
		JN390329.1
	E. decorata	JN390312.1

Appendix 2. Species phylogeny of the four *Euphaea* species reconstructed in BEAST based on 13 protein-coding, two ribosomal, and *A+T-rich* genes of mitochondrial genomes and eight nuclear genes.



Appendix 3. Mutation rates of mitochondrial genes between *Euphaea* species calculated using Tajima-Nei model in MEGA. The mutation rate of 1.77×10^{-8} and 1.2×10^{-8} mutation /site/year of *cox1*.

Gene	*E. y.* vs *E. d.*	*E. y.* vs *E. o.*	*E. f.* vs. *E. d.*	*E. f.* vs *E. o.*	*E. f.* vs *E. y.*	*E. d.* vs *E. o.*	Average
atp8	0.3659	0.3312	0.0114	0.0114	0.3475	0.0231	0.1818 ± 0.1827
atp6	0.3369	0.2513	0.2794	0.1980	0.1061	0.1095	0.2135 ± 0.0934
cox1	0.2309	0.2279	0.2429	0.2443	0.0967	0.0185	0.1769 ± 0.0958
cox2	0.1606	0.1514	0.1389	0.1300	0.1037	0.0130	0.1163 ± 0.0543
cox3	0.2567	0.2713	0.2026	0.2195	0.0998	0.0182	0.1780 ± 0.0989
cob	0.2087	0.2027	0.1151	0.1240	0.1347	0.0222	0.1346 ± 0.0682
nad1	0.2017	0.2158	0.1958	0.1954	0.1013	0.0113	0.1535 ± 0.0809
nad2	0.3370	0.3599	0.3457	0.3578	0.1031	0.0273	0.2552 ± 0.1493
nad3	0.2923	0.2919	0.3199	0.2906	0.0855	0.0311	0.2185 ± 0.1258
nad4	0.1983	0.2067	0.2181	0.2146	0.0735	0.0255	0.1561 ± 0.0842
nad4l	0.1777	0.1622	0.1694	0.1697	0.0702	0.0123	0.1269 ± 0.0690
nad5	0.2300	0.2226	0.2149	0.2184	0.0983	0.0240	0.1680 ± 0.0862
nad6	0.2509	0.2513	0.2509	0.2513	0.0631	0.0108	0.1797 ± 0.1118
s-rRNA	0.1099	0.1022	0.1101	0.1023	0.0093	0.0117	0.0743 ± 0.0495
l-rRNA	0.1250	0.1252	0.1295	0.1297	0.0352	0.0097	0.0924 ± 0.0548
A+T-rich	0.4627	0.5097	0.4696	0.5144	0.0977	0.0739	0.3547 ± 0.2094
Average	0.2466 ± 0.0936	0.2427 ± 0.0996	0.2134 ± 0.1087	0.2107 ± 0.1147	0.1016 ± 0.0722	0.0276 ± 0.0267	0.1738 ± 0.0670

Appendix 4. The organization of mitochondrial genomes of four *Euphaea* species. trn: transfer RNA labelled by the one-letter amino acid code; inc: intergenic nucleotide, negative inc values are overlapping nucleotide sequences of different genes; s1-s4: intergenic spacers; l-rRNA: large subunit of ribosomal gene; s-rRNA: small subunit of ribosomal gene. The incomplete stop codons are labeled in parentheses.

	Strand	Position				Length				Start	Stop	Inc				A+T (%)			
		E. d.	*E. o.*	*E. f.*	*E. y.*	*E. d.*	*E. o.*	*E. f.*	*E. y.*			*E. d.*	*E. o.*	*E. f.*	*E. y.*	*E. d.*	*E. o.*	*E. f.*	*E. y.*
trnI	+	1..67	1..67	1..66	1..66	67	67	66	66			3	3	3	3	65.7	65.7	62.1	62.1
trnQ	-	71..138	71..138	70..137	70..137	68	68	68	68			0	0	-1	-1	72.1	70.6	70.5	67.6
trnM	+	139..207	139..207	137..205	137..205	69	69	69	69			6	6	6	6	72.5	73.9	69.5	72.5
nad2	+	214..1203	214..1203	212..1201	212..1201	990	990	990	990	ATA	TAA	-2	-2	1	1	75.6	74.8	72.3	73.8
trnW	+	1202..1271	1202..1271	1203..1272	1203..1272	70	70	70	70			0	0	0	0	82.9	82.9	75.7	75.7
s1		1272..1287	1272..1287	1273..1290	1273..1290	16	16	18	18			0	0	0	0	68.7	68.7	66.7	72.2
trnC	-	1288..1351	1288..1351	1291..1354	1291..1354	64	64	64	64			0	0	0	0	71.9	71.9	68.7	70.3
trnY	-	1352..1423	1352..1423	1355..1425	1355..1425	72	72	71	71			0	0	0	0	69.4	70.8	64.8	66.2
s2		1424..1458	1424..1458	1426..1460	1426..1460	35	35	35	35			0	0	0	0	85.7	85.7	80	82.9
cox1	+	1459..3006	1459..3006	1461..3008	1461..3008	1548	1548	1548	1548	ATA	TAA	-5	-5	-5	-5	64.8	64.6	64.1	63.8
trnL1	+	3002..3073	3002..3073	3004..3074	3004..3074	72	72	71	71			0	0	0	0	66.7	65.3	63.4	63.4
cox2	+	3074..3761	3074..3761	3075..3762	3075..3762	688	688	688	688	ATG	T(aa)	0	0	0	0	67.7	68.5	67	67.9
trnK	+	3762..3833	3762..3833	3763..3834	3763..3834	72	72	72	72			-1	-1	-1	-1	66.7	66.7	65.3	66.7
trnD	+	3833..3898	3833..3898	3834..3900	3834..3900	66	66	67	67			0	0	0	0	84.8	84.8	83.6	85.1
atp8	+	3899..4057	3899..4057	3901..4059	3901..4059	159	159	159	159	ATC	TAA	-4	-4	-4	-4	76.7	78	76.7	77.4
atp6	+	4054..4728	4054..4728	4056..4730	4056..4730	675	675	675	675	ATA	TAA	-1	-1	-1	-1	69	70.1	70	70.4
cox3	+	4728..5514	4728..5514	4730..5516	4730..5516	787	787	787	787	ATG	T(aa)	0	0	0	0	64.8	64	64.7	64.9
trnG	+	5515..5583	5515..5583	5517..5584	5517..5584	69	69	68	68			0	0	0	0	73.9	72.5	77.9	79.4
nad3	+	5584..5937	5584..5937	5585..5938	5585..5938	354	354	354	354	TTG	TAA	-1	-1	-1	-1	70.3	70.3	71.7	69.8
trnA	+	5937..6004	5937..6004	5938..6005	5938..6005	68	68	68	68			-1	-1	-1	-1	73.5	73.5	72	69.1
trnR	+	6004..6067	6004..6067	6005..6068	6005..6068	64	64	64	64			2	2	2	2	73.4	73.4	73.4	73.4
trnN	+	6070..6136	6070..6136	6071..6137	6071..6136	67	67	67	66			-1	-1	-1	-1	74.6	74.6	73.1	72.7
trnS1	+	6136..6206	6136..6206	6137..6207	6136..6206	71	71	71	71			1	1	1	1	63.4	64.8	64.8	64.8
trnE	+	6208..6274	6208..6274	6209..6275	6208..6274	67	67	67	67			-2	-2	-2	-2	82.1	82.1	82.1	80.6
trnF	-	6273..6340	6273..6340	6274..6342	6273..6341	68	68	69	69			2	2	2	2	70.6	72.1	71	71
nad5	-	6343..8065	6343..8065	6345..8067	6344..8066	1723	1723	1723	1723	ATT	T(aa)	0	0	1	0	69.7	69.8	70.6	71
trnH	-	8066..8131	8066..8131	8069..8133	8067..8132	66	66	65	66			0	0	2	2	69.7	68.2	67.7	69.7
nad4	-	8132..9475	8132..9475	8136..9497	8135..9498	1344	1344	1344	1344	ATG	TAA	-7	-7	-7	-7	71.1	71.5	72	71.9
nad4L	-	9469..9762	9469..9762	9473..9766	9472..9765	294	294	294	294	ATG	TAA	2	2	2	2	73.1	73.1	74.1	73.1
trnT	+	9765..9832	9765..9832	9769..9835	9768..9834	68	68	67	67			0	0	0	0	75	75	73.2	71.6

s3		9833..9844	9833..9845	9836..9848	9835..9848	12	13	13	14			0	0	0	0	41.7	38.5	38.5	50
trnP	-	9845..9910	9846..9911	9849..9914	9849..9914	66	66	66	66			1	1	1	1	77.3	78.8	77.3	77.3
nad6	+	9912..10409	9913..10410	9916..10413	9916..10413	498	498	498	498	ATC	TAA	-1	-1	-1	-1	72.5	72.3	73.5	73.5
cob	+	10409..11542	10410..11543	10413..11546	10413..11546	1134	1134	1134	1134	ATG	TAA	-2	-2	-2	-2	68.5	68.5	67.7	69.1
trnS2	+	11541..11604	11542..11605	11545..11608	11545..11608	64	64	64	64			0	0	0	0	76.6	76.6	68.7	76.6
s4		11605..11620	11606..11621	11609..11624	11609..11624	16	16	16	16			0	0	0	0	81.2	81.2	75	81.2
nad1	-	11621..12571	11622..12572	11625..12575	11625..12575	951	951	951	951	TTG	TAG	1	1	1	1	68.9	69	69.5	69.1
trnL2	-	12573..12641	12574..12641	12577..12643	12577..12643	69	68	67	67			0	0	0	0	76.8	77.9	73.1	73.1
l-rRNA	-	12642..13930	12642..13930	12644..13931	12644..13931	1289	1289	1288	1288			0	0	0	0	64.8	64.8	73.9	72.2
trnV	-	13931..14001	13931..14001	13932..14003	13932..14003	71	71	72	72			0	0	0	0	70.1	70.5	70.9	71
s-rRNA	-	14002..14776	14002..14776	14004..14781	14004..14781	775	775	778	778			0	0	0	0	75	74.9	70.8	74.7
A+T-rich		14777..15861	14777..15863	14782..15700	14782..15709	1085	1087	919	928			0	0	0	0	82.9	83.2	80.3	80.1

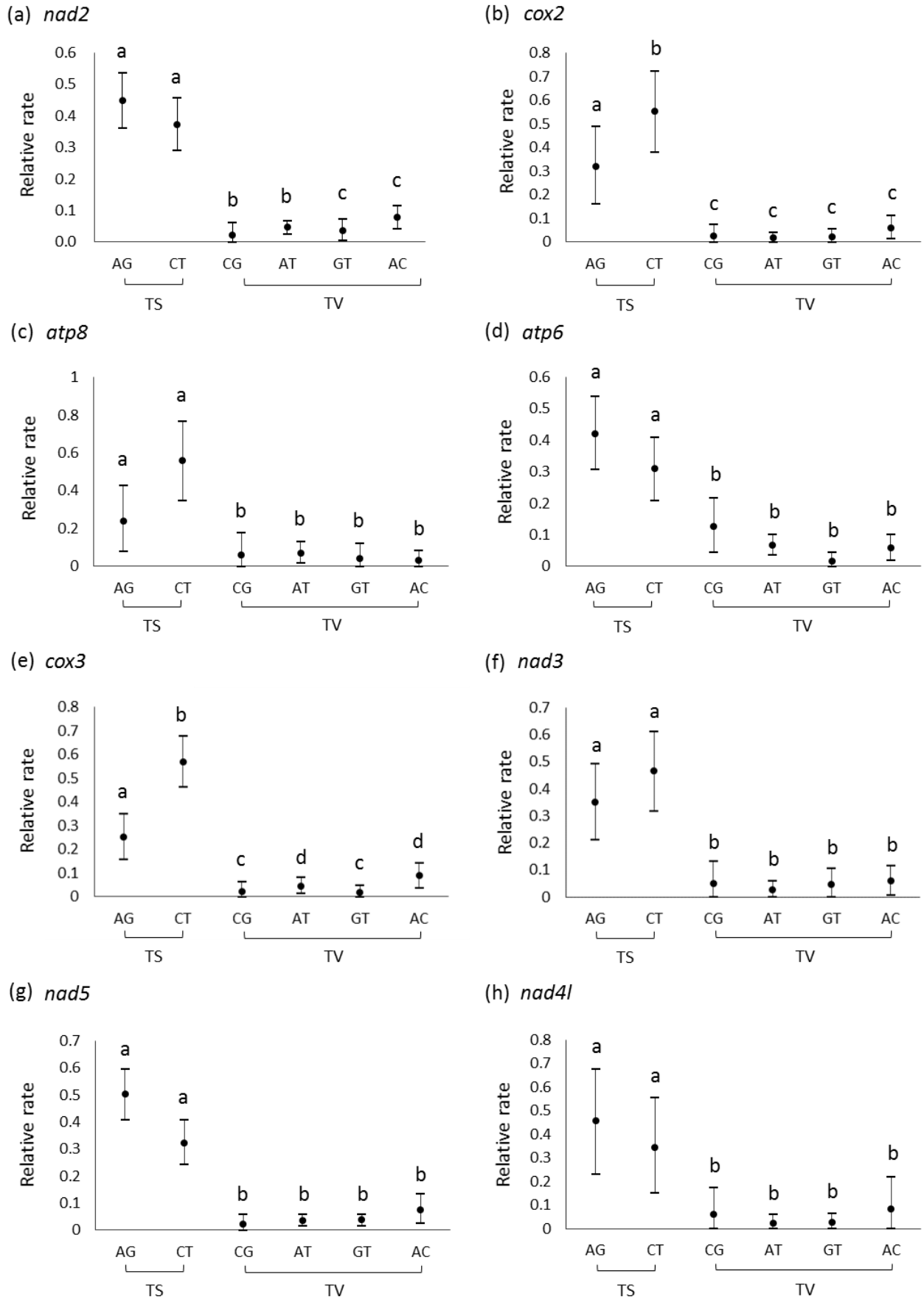
Appendix 5. The d_N/d_S ratio (ω) of mitochondrial and nuclear genes calculated using the branch model in codeml option of PAML. Model 0 has only one ω ratio (ω_T) for all four *Euphaea* species; model 2 assumes each species has different ω ratios (ω_d, ω_o, ω_f and ω_y for ω ration in *E. decorata*, *E. ornata*, *E. formosa* and *E. yayeyamana*, respectively). N is the total number of haplotypes used. Numbers in bold indicate positive selection.

Level		N	Length	df	Likelihood of model		χ^2	p	ω_T	ω_y	ω_f	ω_o	ω_d
					m0	m2							
Population	mtDNA												
	cox2	32	498	4	-1478.3777	-1475.3633	6.0288	0.1970	0.0132				
	nad5	38	909	4	-2194.8263	-2183.0809	23.4909	0.0001	-	0.0085	0.0985	0.0663	0.0123
	nuDNA												
	act	14	423	4	-652.6245	-651.1693	2.9103	0.5730	0.0400				
	arr	22	303	4	-796.9831	-793.5623	6.8416	0.1445	0.1052				
	awd2	10	153	4	-236.2715	-236.2715	$2x10^{-6}$	1.0000	**999.000** (d_N: 0.0564; d_S< 0.0001)				
	EF1α	15	705	4	-1139.0231	-1133.7889	10.4685	0.0332	-	0.0828	0.8305	**1.2502**	0.2070
	fer	21	333	4	-701.9700	-701.1698	1.6006	0.8087	0.0536				
	mlc	10	162	4	-274.3832	-272.7528	3.2607	0.5152	0.1158				
	opsin	16	309	4	-524.4279	-521.7102	5.4355	0.2455	0.1227				
	sdhB	15	258	4	-527.8787	-524.6821	6.3931	0.1717	0.1719				
	Average								0.1015				
Species	mtDNA												
	atp6		672	4	-1358.9059	-1329.9486	57.9147	< 0.0001	-	0.0422	0.0228	0.0001	0.5742
	atp8		156	4	-247.6631	-247.4845	0.3572	0.9858	0.0186				
	cob		1131	4	-1953.7438	-1951.2565	4.9747	0.2899	0.0095				
	cox1		1545	4	-2721.4808	-2717.5284	7.9047	0.0951	0.0044				
	cox2		687	4	-1147.5413	-1147.2068	0.6690	0.9551	0.0124				
	cox3		786	4	-1388.6472	-1385.0979	7.0986	0.1308	0.0109				
	nad1		948	4	-1623.8951	-1619.6706	8.4490	0.0764	0.0137				
	nad2		987	4	-1916.8673	-1913.2311	7.2725	0.1222	0.0275				
	nad3		348	4	-677.8625	-676.2565	3.2121	0.5230	**11.6780** (d_N: 0.4461; d_S: 0.0382)				
	nad4		1341	4	-2240.4020	-2236.7339	7.3361	0.1192	0.0107				
	nad4l		291	4	-460.2535	-459.3542	1.7985	0.7728	0.0128				
	nad5		1722	4	-2980.5782	-2980.2613	0.6338	0.9592	0.0087				
	nad6		495	4	-900.4672	-899.6510	1.6325	0.8029	0.0548				
	Average								0.0167				

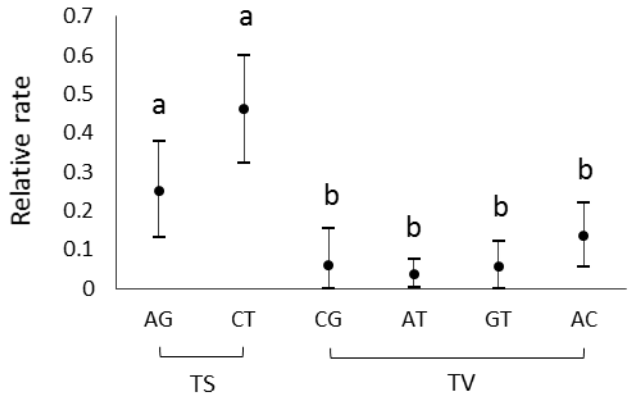
Appendix 6. The d_N and d_S ratio of the gene sequences with different ω in the four *Euphaea* species.

Gene name	*E. decorata*	*E. ornata*	*E. formosa*	*E. yayeyamana*
	ω (d_N, d_S)	ω (d_N, d_S)	ω (d_N, d_S)	ω (d_N, d_S)
mtDNA				
atp6	0.0422 (0.046464, 0.080927)	0.0228 (0.000007, 0.065363)	0.0001 (0.007407, 0.324724)	0.5741 (0.001805, 0.042812/0
nad5	0.0085 (0.000781, 0.063669)	0.0985 (0.001447, 0.02182)	0.0663 (0.001403, 0.014247)	0.01239 (0.015353, 2.745501)
nuDNA				
EF1α	0.0828 (0.000003, 0.000001)	0.8305 (0.000001, 0.000001)	1.2502 (0.000001, 0.000002)	0.2070 (0.00034, 0.004102)

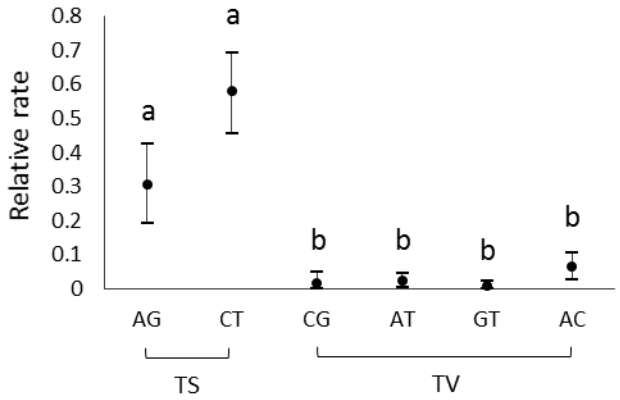
Appendix 7. The transformation rate matrix of mitochondrial and nuclear genes (mean ± 95% credibility interval, CI). The alphabet above the bars represents the rate categories estimated in model selection of MEGA.



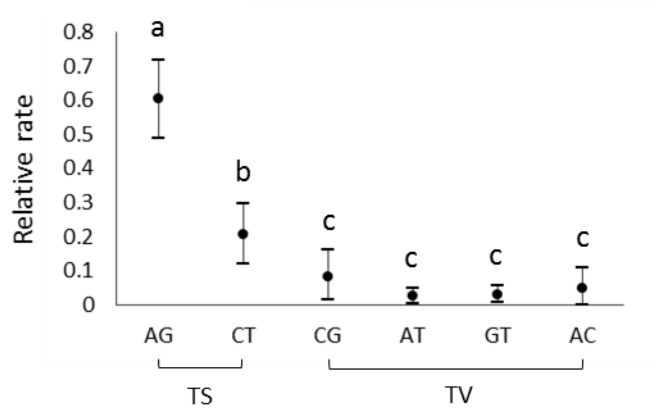
(i) *nad6*



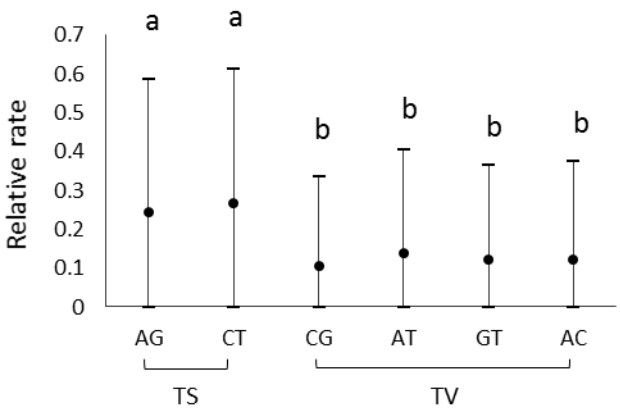
(j) *cob*



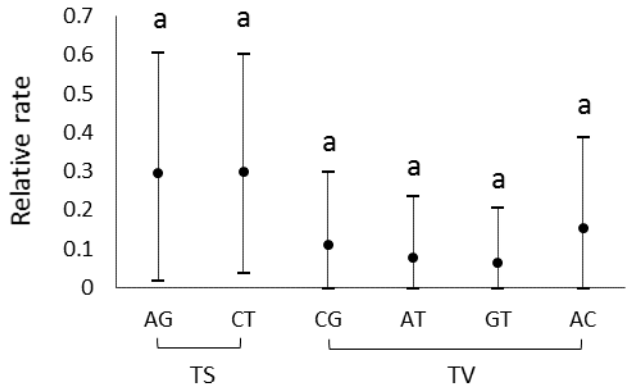
(k) *nad1*



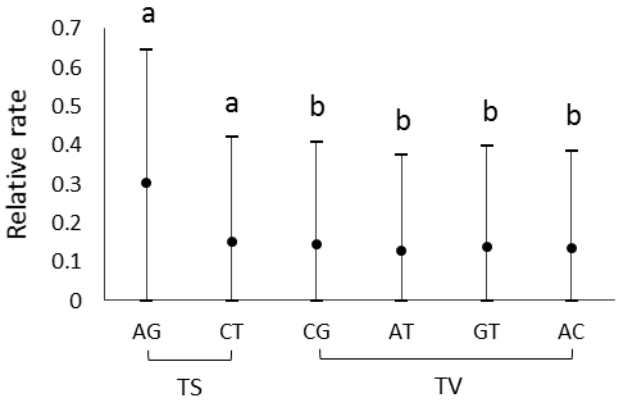
(l) *act*



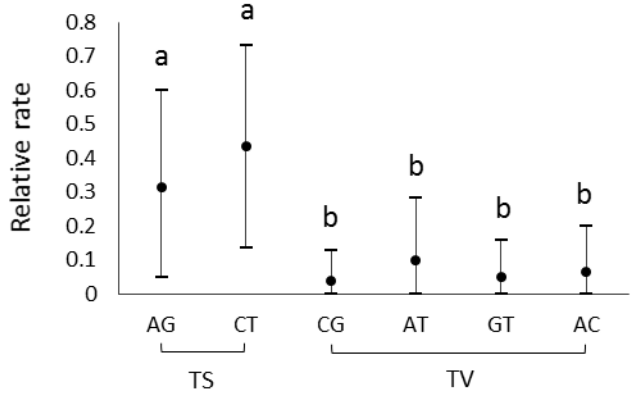
(m) *arr* exon



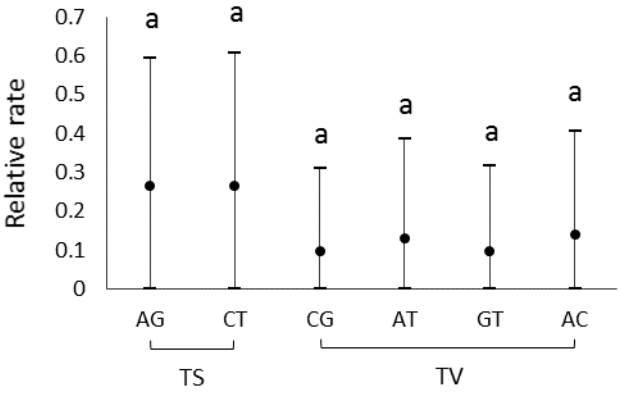
(n) *awd2* exon



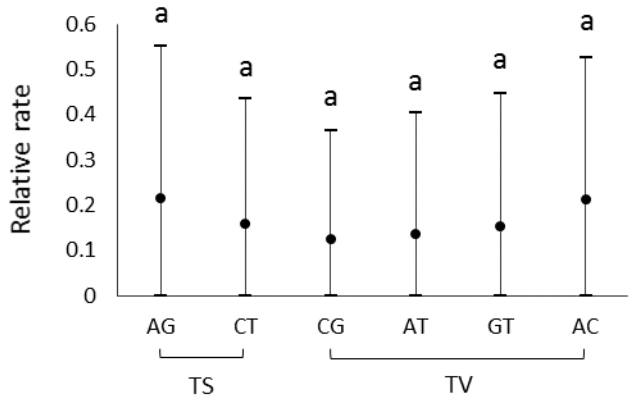
(o) *fer* exon



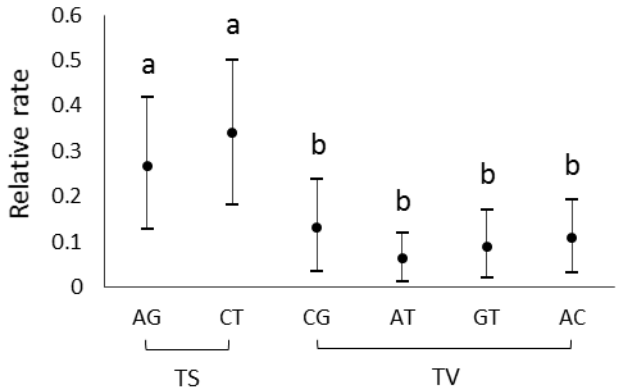
(p) *opsin* exon



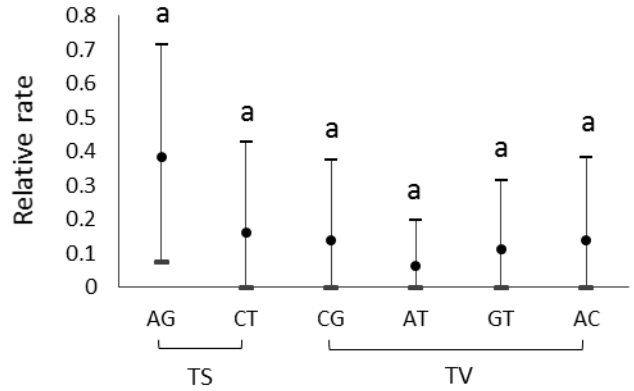
(q) *sdhB* exon



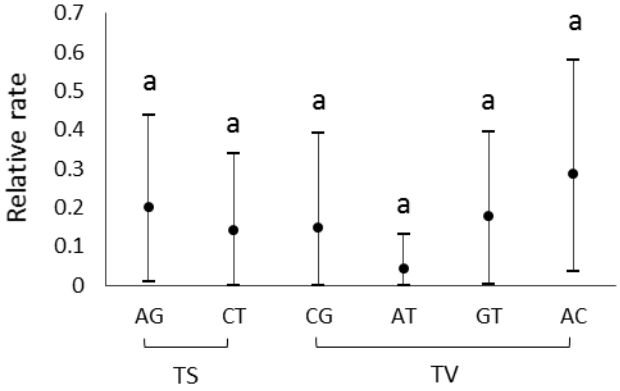
(r) *arr* intron



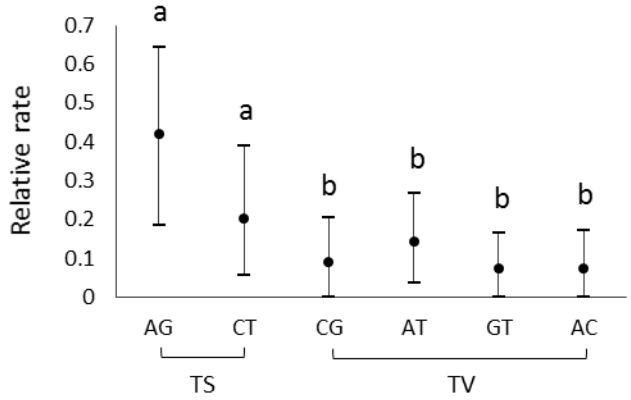
(s) *awd2* intron



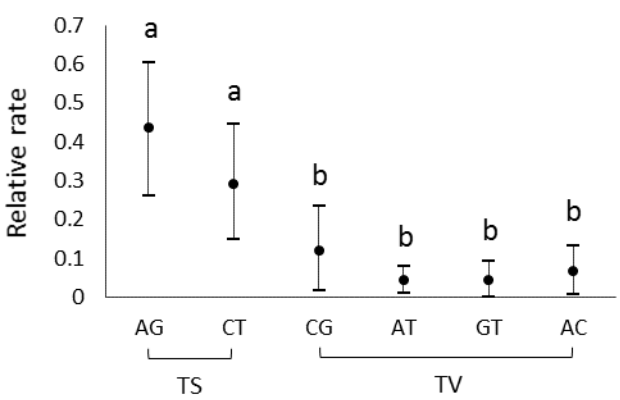
(t) *EF1α* intron



(u) *fer* intron



(v) *mlc* intron



(w) *opsin* intron