Systematic Entomology

Systematic Entomology (2018), **43**, 702–715 DOI: 10.1111/syen.12299

Some mitochondrial genes perform better for damselfly phylogenetics: species- and population-level analyses of four complete mitogenomes of *Euphaea* sibling species

YUN-CHIEH CHENG[1], MING-YU CHEN[1], JO-FAN WANG[2], AI-PING LIANG[3] and CHUNG-PING LIN[2,4]

[1]Department of Life Science & Center for Tropical Ecology and Biodiversity, Tunghai University, Taichung, Taiwan, [2]Department of Life Science, National Taiwan Normal University, Taipei, Taiwan, [3]Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, China and [4]Laboratory of Forest Ecology, Graduate School of Agriculture, Kyoto University, Kyoto, Japan

Abstract. Animal mitochondrial genes continue to provide an efficient and inexpensive assessment of genetic diversity. However, which mitochondrial genes should be selected to best estimate species phylogeny and population genealogy remains uncertain for most under-sampled taxa. We analysed four complete mitochondrial genomes of sibling species of *Euphaea* damselflies, *E. decorata*, *E. ornata*, *E. formosa* and *E. yayeyamana* (Insecta, Odonata, Euphaeidae), to examine the patterns of selection and to evaluate the phylogenetic utility of the mitochondrial genes compared with nuclear genes. The results indicated that mitochondrial protein-coding *nad2* (NADH dehydrogenase subunit 2) and noncoding *A + T-rich* (control region) genes have the highest mutation rates and more phylogenetic utility [higher parsimony-informative sites; higher α (the shape parameter of gamma distribution); lower rates of heterogeneity among sites; and higher relative substitution rates] than all the other mitochondrial and nuclear genes analysed. In contrast, the animal DNA barcoding gene cytochrome *c* oxidase subunit 1 (*cox1*) had average values for all estimated parameters of phylogenetic performance and was sometimes outperformed by other mitochondrial genes. The majority of the mitochondrial and nuclear genes in *Euphaea* damselflies have experienced frequent purifying selection, except for two cases of potential positive selection in NADH dehydrogenase subunit 3 (*nad3*) and elongation factor 1α (*EF1*α), and all mitochondrial genes had experienced stronger purifying selection than nuclear genes. Our findings indicated that mitochondrial *nad2* and the *A + T-rich* region should be selected to provide efficient and high-resolution phylogenetic markers for damselflies at the species and population level.

Introduction

Mitochondrial genes are the most extensively used genetic markers in animal phylogenetics and population genetics (Boore 1999; Simon *et al.* 2006). They continue to provide an efficient and inexpensive measure of genetic diversity at the species and population levels, such as with DNA barcodes (Hebert *et al.* 2003) and other applications (e.g. Miraldo *et al.*, 2016). For insects, individual mitochondrial genes are popular for use as genetic markers and have an advantage over other genetic markers because of their low-cost, nonrecombination, availability of universal primers, ease of amplification, possession of both conserved and variable regions, and smaller effective population sizes for fast reciprocal monophyly (Sunnucks 2000; Avise 2004; Rubinoff & Holland 2005; Simon *et al.* 2006). However, because mitochondrial DNA (mtDNA) is inherited as a unit, the use of single mitochondrial genes for phylogenetic

Correspondence: Chung-Ping Lin, No. 88, Section 4, Tingzhou Road, Department of Life Science, National Taiwan Normal University, Taipei 11677, Taiwan. E-mail: treehopper@ntnu.edu.tw

inference has several well-known disadvantages when used alone, including failure to detect interspecific hybridization, incomplete lineage sorting and mitochondrial introgression (Funk & Omland 2003; Avise 2004; Rubinoff & Holland 2005; Galtier *et al.* 2009). These disadvantages can be turned into advantages when mtDNA is used in combination with independent genetic markers. Earlier studies have also suggested that mitochondrial genes evolve too rapidly and show nucleotide substitution patterns that are problematic for deeper-level insect phylogenetics (e.g. Lin & Danforth, 2004). Compared with nuclear genes, mitochondrial genes showed higher inconsistencies among phylogenetic signals (lower values of consistency index, CI) and less symmetrical transformation rate matrices of nucleotide substitutions, suggesting that mitochondrial genes are lower-quality genes for phylogenetic analysis at higher taxonomic levels (Lin & Danforth 2004; Caravas & Friedrich 2013). However, these studies analysed deeper divergence (Mesozoic and older) of higher-level (generic, tribal) phylogenetic relationships of insects and evaluated only the most commonly used mitochondrial genes such as cytochrome *c* oxidase subunit 1 (*cox1*) and subunit 2 (*cox2*). The phylogenetic utility of other mitochondrial genes and the relative utility of mitochondrial genes at shallow phylogenetic divergence (Pleistocene and younger) of lower taxonomic level (species, population) phylogenetic relationships are poorly understood, especially for less frequently utilized genes (Simon *et al.* 1994, 2006).

Nearly evolutionarily neutral and nonrecombining mitochondrial genes have been the most widely used markers to study population history and diversity because it was believed that mitochondrial diversity should reflect the effective population size (Frankham 1996; Avise 2004). Despite their popularity, studies have suggested that the genetic diversity of mitochondrial genes might not be positively correlated with the population size, which challenges their utility in population genetic research (Bazin *et al.* 2006; Galtier *et al.* 2009; Piganeau & Eyre-Walker 2009; Stoeckle & Thaler 2014). One hypothesized mechanism for the decrease in mitochondrial diversity in larger populations is through "genetic draft" (Meiklejohn *et al.* 2007), which is a process of fixing advantageous mutations and reducing mitochondrial diversity by frequent positive selection and hitchhiking (Neher 2013). Because the effect of genetic draft is much stronger in larger populations, invertebrates with larger population sizes, such as insects, tend to have lower mitochondrial genetic diversity than vertebrates with smaller populations (Bazin *et al.* 2006; Meiklejohn *et al.* 2007). However, the level and frequency of positive selection and genetic draft acting on insect mitochondrial genomes are still largely unknown (but see Roux *et al.* 2014).

With the advancement of more efficient next-generation sequencing technologies (Goodwin *et al.* 2016), sequencing and analysing the entire mitochondrial genome for insect phylogenetics has become more popular (reviewed in Cameron 2014; Misof *et al.* 2014; López-López & Vogler 2017). However, sequencing the entire mitochondrial genome is not always feasible due to financial and technological limitations, nor always desirable for addressing particular biological questions and taxonomic levels (McDonagh *et al.* 2016; Chesters 2017). Therefore, single mitochondrial genes continue to provide a reliable, fast and inexpensive alternative to whole-genome sequencing. Some of the most interesting biological questions in insect ecology and evolution lie at the boundary of species and populations (Grimaldi & Engel 2005; Chesters 2017). Therefore, it is essential to identify appropriate genetic markers with high resolution at the species/population level. Choosing genetic markers is the first important decision in all successful molecular phylogenetic and phylogeographic studies of closely related organisms (Sunnucks 2000). But which mitochondrial genes should be selected to best estimate species phylogeny or population genealogy within a group of insects? The analysis of mitogenomes of congeneric species provides a promising approach for the assessment of evolutionary trends of mitochondrial genes at short divergence times in under-sampled taxa (Gissi *et al.* 2008).

We addressed this question by analysing the complete mitochondrial genomes of four *Euphaea* Selys damselflies (Odonata, Euphaeidae): *E. decorata* (widespread in Southeast Asia), *E. ornata* (Hainan Island), *E. formosa* (Taiwan) and *E. yayeyamana* (Yaeyama Islands of Japan). A molecular phylogeny of *Euphaea* indicated that these four species consist of two sibling species pairs (Lee & Lin 2012a); thus, they provide a suitable case study for examining the phylogenetic performance of mitochondrial genes at the species and population levels in damselflies. Eight nuclear genes of the same four species were used as comparisons (Lee & Lin 2012a). First, we examined the level of selection in mitochondrial and nuclear genes and predicted that the nuclear genes exhibit a higher level of selection than the mitochondrial genes (Meiklejohn *et al.* 2007) because the effective population size of nuclear genes is, in general, four times larger than that of mitochondrial genes (Ballard & Whitlock 2004). Second, we compared the phylogenetic utility of the mitochondrial genes and the nuclear genes by calculating the parameters of nucleotide substitution and inferring the patterns of rate variation among genes. Specifically, we examined the differences in substitution rate among mitochondrial and nuclear genes and tested for positive relationships between base composition (A + T bias) and relative rate, between CI and the shape parameter of gamma distribution (α), parsimony-informative (PI) sites and relative rate, and between proportion of invariant sites (*Pi*) and α (Lin & Danforth 2004).

Materials and methods

Mitochondrial genome sequencing

The complete mitochondrial genomes of three *Euphaea* sibling species (*E. decorata*, *E. ornata* and *E. yayeyamana*) were sequenced, and a fourth mitochondrial genome sequence of *E. formosa* was obtained from Lin *et al.* (2010). The specimens of *E. decorata*, *E. ornata* and *E. yayeyamana* used for genome sequencing were collected in Hong Kong, Hainan Island and the Yaeyama Islands, respectively (Huang & Lin 2011; Lee & Lin 2012a, b). The collected specimens were placed in 95% ethanol and then stored in −80°C freezers. Total genomic DNA was extracted using a standard protocol

(Lee & Lin 2012a). Two sets of long PCR primers (Yamauchi *et al.*, 2004) were used to amplify the complete mitochondrial genome. DNA sequencing was completed by primer walking using 22 *Euphaea*-specific mitochondrial primers (Lin *et al.* 2010). The genome sequences (GenBank accession numbers: *E. yayeyamana*, KF718293; *E. decorata*, KF718294; *E. ornata*, KF718295) were edited and assembled following an established procedure (Lin *et al.* 2010) and annotated in reference to the mitochondrial genome of *E. formosa* (HM126547). The mitochondrial genome of *Euphaea* contains the entire set of 37 mitochondrial genes [13 protein-coding genes (ATPase subunits 6 and 8, *atp6* and *atp8*; NADH dehydrogenase subunits 1–6, 4 L, *nad1–6*, *4 L*; cytochrome b, *cob*; cytochrome *c* oxidase subunits 1–3, *cox1–3*), 22 tRNA genes, two rRNA genes (large and small subunits of ribosomal gene, l-rRNA and s-rRNA, respectively), and the *A + T-rich* control region]. A total of 75 sequences of two mitochondrial genes (*cox2*, *nad5*) and 136 sequences of eight nuclear genes (actin, *act*; arrestin2, *arr*; abnormal wing discs 2, *awd2*; elongation factor 1α, *EF1α*; *ferritin*, *fer*; myosin light chain, *mlc*, long-wavelength opsin, *opsin*; succinate dehydrogenase B, *sdhB*) from *Euphaea* populations were obtained from published studies (Appendix S1). Additionally, one *cox2* gene sequence (specimen code: EfAa12wi), eight *nad5* gene sequences (EfAa1, EfAa2, EfBa1, EfBa5, EfBa12, EfDa4, EfTa3, TW1) and one *EF1α* (EfAa9) gene sequence from *E. formosa* were sequenced. The nucleotide sequences of protein-coding genes were translated into amino acid sequences using invertebrate mitochondrial genetic codes in the SIB ExPASy Bioinformatics Resource Portal (Gasteiger *et al.* 2003) to confirm reading frames and identify potential sequencing errors. The haplotype numbers of gene sequences were inferred using dnaSP v5 (Librado & Rozas 2009). For the population-level analysis, we analysed 137 nuclear sequences and 84 mitochondrial sequences of 10 genes (mitochondrial: *cox2*, *nad5*; nuclear: *act*, *arr*, *awd2*, *EF1α*, *fer*, *mlc*, *opsin*, *sdhB*) from multiple individuals of the four species (Appendix S1). For the species-level analyses, we randomly used one sequence of the 13 mitochondrial protein-coding genes, two noncoding genes, the *A + T-rich* region, and eight nuclear genes from each of the four *Euphaea* species.

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

The d_N/d_S ratio (ω) of the gene sequences was calculated to test for signatures of natural selection using the branch model of the codeml option in PAML v.4.7 (Yang 2007). A value of ω significantly greater than 1 indicates positive selection, whereas a value of ω significantly less than 1 suggests purifying selection. Extremely large and small values of ω (> 100 or < 0.001) were excluded from the analysis because these values were probably due to an insufficient number of substitutions. The best nucleotide substitution models selected in MEGA 5.2 using Bayesian information criterion were as follows: 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*; and TN93 – *cox3*. Two models were used to compare the strengths of selection using maximum likelihood: model = 0, where there is only one ω ratio for all four species (ω_T, total ω); and model = 2, where each species has its own ω ratio (ω_d, ω_o, ω_f and ω_y, the ω of *E. decorata*, *E. ornata*, *E. formosa* and *E. yayeyamana*, respectively). The selection strengths were different among species if the null hypothesis (model = 0) was rejected. The Hyphy program of MEGA 6.0 (Tamura *et al.* 2013) was used to calculate the level of selection across codon sites within individual genes.

Phylogenetic analysis and nucleotide substitution patterns

A total of 24 genes (13 protein-coding, three noncoding mitochondrial genes and eight nuclear genes) were used to estimate the species tree of four *Euphaea* species in *BEAST v.1.8.2 (Drummond *et al.* 2012). The best substitution models from the PAML analyses were used, with additional models for *s-rRNA*, *l-rRNA* (HKY) and the *A + T-rich* region (HKY + G). The Markov chain Monte Carlo (MCMC) procedure was executed for 1×10^8 generations, with trees sampled every 1×10^4 generations and the initial 2×10^7 generations discarded as burn-in. The convergence of MCMC runs was assessed using the values of the effective sample size (ESS > 741) in TRACER v.1.6 (Rambaut *et al.* 2014).

The pairwise genetic distances of the haplotypes were calculated in MEGA 6.0 (Tamura *et al.* 2013). The number of substitutions per site within and between species was calculated using the Tajima–Nei model (Tajima & Nei 1984). The pairwise genetic distance was calibrated using the *cox1* mutation rate of 1.77×10^{-8} mutations/year per site (Papadopoulou *et al.*, 2010) for all mitochondrial genes, and a substitution rate of 1.2×10^{-8} mutations/year per site from the nuclear gene *arr* for nuclear genes (Moriyama & Gojobori 1992). The PI sites were calculated in MEGA 6.0.

Two nucleotide substitution models, GTR + SSR (site-specific rates) and GTR + I + G (invariant sites and gamma distribution), were used to calculate the relative substitution rates among genes in MRBAYES v.3.2 (Huelsenbeck & Ronquist 2001). The gene partitions included 13 mitochondrial protein-coding genes, three noncoding regions (*s-rRNA*, *l-rRNA* and the *A + T-rich*) and eight nuclear genes. For site-specific rates (SSRs), we partitioned the substitution rates of protein-coding genes into three codon sites for comparison. In GTR + SSR analyses, the MCMC was run for 1×10^7 generations, with trees sampled every 1000 generations. For the GTR + I + G model, the *Q* matrix (transformation rate matrix), α of gamma distribution and *Pi* were obtained to estimate the relative symmetry of substitution and the level of rate heterogeneity among sites. The α value estimated the rate heterogeneity within the genes; a higher α value suggests lower heterogeneity of substitution rates among sites. In the GTR + I + G analysis, the MCMC was run for 3×10^7 generations with a sampling frequency of every 1000 iterations. The convergence of all MCMC runs was confirmed when the standard deviation of each split frequency was below 0.01 and the potential scale reduction factor was equal to 1. The correlation between parameter estimates was tested using linear regression.

Table 1. The organization of the mitochondrial genomes of four *Euphaea* species

		Position				Length						Inc				A + T (%)			
	Strand	*E. d.*	*E. o.*	*E. f.*	*E. y.*	*E. d.*	*E. o.*	*E. f.*	*E. y.*	Start	Stop	*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..10 409	9913..10 410	9916..10 413	9916..10 413	498	498	498	498	ATC	TAA	−1	−1	−1	−1	72.5	72.3	73.5	73.5
cob	+	10 409..11 542	10 410..11 543	10 413..11 546	10 413..11 546	1134	1134	1134	1134	ATG	TAA	−2	−2	−2	−2	68.5	68.5	67.7	69.1
trnS2	+	11 541..11 604	11 542..11 605	11 545..11 608	11 545..11 608	64	64	64	64			0	0	0	0	76.6	76.6	68.7	76.6
s4		11 605..11 620	11 606..11 621	11 609..11 624	11 609..11 624	16	16	16	16			0	0	0	0	81.2	81.2	75	81.2
nad1	−	11 621..12 571	11 622..12 572	11 625..12 575	11 625..12 575	951	951	951	951	TTG	TAG	1	1	1	1	68.9	69	69.5	69.1
trnL2	−	12 573..12 641	12 574..12 641	12 577..12 643	12 577..12 643	69	68	67	67			0	0	0	0	76.8	77.9	73.1	73.1
l-rRNA	−	12 642..13 930	12 642..13 930	12 644..13 931	12 644..13 931	1289	1289	1288	1288			0	0	0	0	64.8	64.8	73.9	72.2
trnV	−	13 931..14 001	13 931..14 001	13 932..14 003	13 932..14 003	71	71	72	72			0	0	0	0	70.1	70.5	70.9	71
s-rRNA	−	14 002..14 776	14 002..14 776	14 004..14 781	14 004..14 781	775	775	778	778			0	0	0	0	75	74.9	70.8	74.7
A + T-rich		14 777..15 861	14 777..15 863	14 782..15 700	14 782..15 709	1085	1087	919	928			0	0	0	0	82.9	83.2	80.3	80.1

Trn, transfer RNA labelled by the one-letter amino acid code; *inc*, intergenic nucleotide, negative inc values represent overlapping nucleotide sequences of different genes; *s1-s4*, intergenic spacers; *l-rRNA*, large subunit of ribosomal gene; *s-rRNA*, small subunit of ribosomal gene: *E. d.*, *Euphaea decorata*; *E. o.*, *Euphaea ornata*; *E. f.*, *Euphaea formosa*; *E. y.*, *Euphaea yayeyamana*. The incomplete stop codons are labelled in parentheses.

Data accessibility

The assembled mitogenomes of the four *Euphaea* species are available at GenBank (accession numbers: *E. decorata*, KF718294; *E. ornata*, KF718295; *E. formosa*, HM126547; *E. yayeyamana*, KF718293).

Results

Mitochondrial genomes

The mitochondrial genome sizes of the sister species pair, *E. decorata* (15 861 bp) and *E. ornata* (15 863 bp), were larger than those of the other sister species pair, *E. formosa* (15 700 bp) and *E. yayeyamana* (15 709 bp) (Table 1), suggesting phylogenetic conservatism in the evolution of mitochondrial genome size. Most genome size variation occurred in the *A + T-rich* region. The gene arrangement and composition of three newly sequenced mitochondrial genomes of *E. decorata*, *E. ornata* and *E. yayeyamana* were the same as those of *E. formosa* (Lin *et al.* 2010).

Phylogeny and sequence divergence

The species tree revealed two pairs of sister taxa, *E. decorata* + *E. ornata* and *E. formosa* + *E. yayeyamana* (Appendix S2), consistent with earlier studies (Huang & Lin, 2011; Lee

& Lin, 2012a, b). The tree branches leading to *E. decorata* and *E. ornata* were shorter than those leading to *E. formosa* and *E. yayeyamana*, suggesting that *E. ornata* and *E. decorata* have diverged more recently than the split between *E. formosa* and *E. yayeyamana*. The pairwise sequence divergences of mitochondrial genes were approximately six times higher than for nuclear genes (mitochondrial, 0.0928 ± 0.0054; nuclear, 0.014 ± 0.0045) (Appendix S3). The *nad2* gene and the *A + T-rich* region had the highest estimated mutation rates among all mitochondrial genes (0.2552 ± 0.1493 and $0.3547 \pm 0.2094 \times 10^{-9}$ mutations/year per site, respectively; Appendix S4).

Patterns of selection

At the population level, the four *Euphaea* species were inferred to have the same ω-values of less than 1 for all examined sequences except for *nad5* and *EF1α* (Appendix S5), indicating that these genes are under similar levels of purifying selection across species. Nuclear *awd2* had an unrealistically large ω_T (999) due to a small d_S value (< 0.0001). The ω-values of *nad5* in the four *Euphaea* species were significantly different ($\chi^2 = 23.49$, $P = 0.0001$), suggesting variable purifying selection across species. The least populous *E. yayeyamana* from the Yaeyama Islands (Lee & Lin 2012a) had the lowest ω (*nad5*, 0.0085), indicating that this species had experienced the strongest purifying selection in *nad5*, whereas *E. formosa* from Taiwan, with a median population size, had the highest ω (*nad5*, 0.0985), suggesting that it had experienced the weakest purifying selection among the four species. The ω of *EF1α* in *E. ornata* from Hainan Island was significantly greater than 1 ($\omega = 1.25$; $d_N = 0.0015$, $d_S = 0.0012$), revealing that this gene might have experienced positive selection in *E. ornata*.

The average ω-values of mitochondrial genes at both the population and species levels were significantly lower than those of nuclear genes at the population level (Fig. 1), suggesting that mitochondrial genes experienced stronger purifying selection than nuclear genes. For species-level analyses of 13 mitochondrial protein-coding genes, the one-ratio model (m0) described the variation of ω better than the four-ratio model (m2), except for *atp6* ($\chi^2 = 57.915$, $P < 0.0001$; Appendix S5). All 13 mitochondrial genes were under purifying selection, except for *nad3*, which showed a signature for positive selection ($\omega_T = 11.6780$) due to a small d_S value ($d_N = 0.4461$; $d_S = 0.0382$; Appendix S6). For *atp6*, the two *Euphaea* species (*E. formosa* and *E. ornata*) with medium population sizes had small ω-values (0.0228 and 0.0001, respectively), whereas the most populous species, *E. decorata*, had a large ω-value (0.5742), and the least populous species *E. yayeyamana* also had a small ω (0.0422). Sliding windows revealed potential positive selection between codon sites 145–151 of *awd2* ($d_N - d_S > 0$; Fig. 2c). Among eight nuclear protein-coding genes (Fig. 2a–h), *EF1α* and *opsin* had experienced much more frequent purifying selection than the others, which had experienced only sporadic purifying selection across sites. A few nuclear codon sites had experienced stronger purifying selection ($d_N - d_S < -10$) than the others

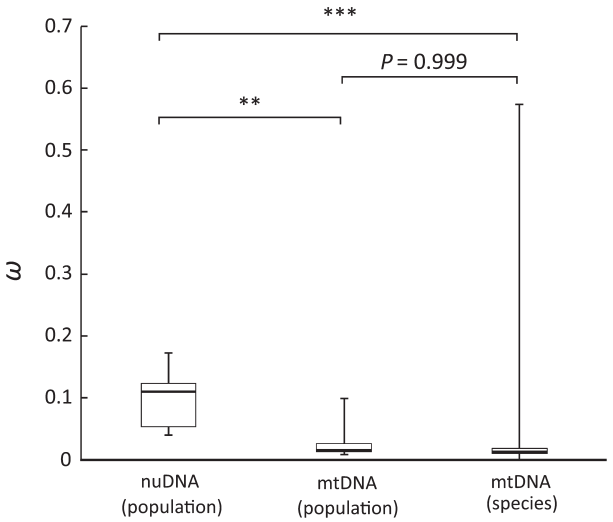

Fig. 1. The ω (d_N/d_S) of mitochondrial (mt) and nuclear (nu) genes in population- and species-level analyses of *Euphaea*. The ω of nuclear genes (0.102 ± 0.045) is significantly different from that of the mitochondrial genes at the population (0.003 ± 0.034) and species (0.029 ± 0.081) levels (Tukey test).

(e.g. codon sites 13 and 151 of *arr*, codon site 10 and 16 of *fer*, and codon site 250 of *sdhB*). Mitochondrial *cox2* and *nad5* had also experienced purifying selection at most codon sites (Fig. 2i, j). Overall, purifying selection occurred more frequently and was more widespread in mitochondrial than nuclear genes.

Parsimony-informative sites and consistency index

The *nad2* and the *A + T-rich* region had the highest number of PI sites among all mitochondrial genes (*nad2* = 0.129, *A + T-rich* = 0.128; Table 2). Among the nuclear genes, the introns of *opsin* and *EF1α* had the highest number of PI sites (*opsin* = 0.057; *EF1α* = 0.056). The average number of PI sites in mitochondrial genes (0.077 ± 0.342) was approximately 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 the species level. Mitochondrial and nuclear genes had overall similar high CIs (mitochondrial, 0.980 ± 0.023, nuclear, 0.965 ± 0.091, $t = 0.561$, $p = 0.290$; Table 2), suggesting that they had low amounts of homoplasy.

Relative substitution rates

For both mitochondrial and nuclear genes, the third codons had the highest substitution rates, followed by the first and second codons, except for the second codons of *awd2*, which had the highest rate among the three codons (Fig. 3a). The relative substitution rates of nuclear genes were significantly lower than

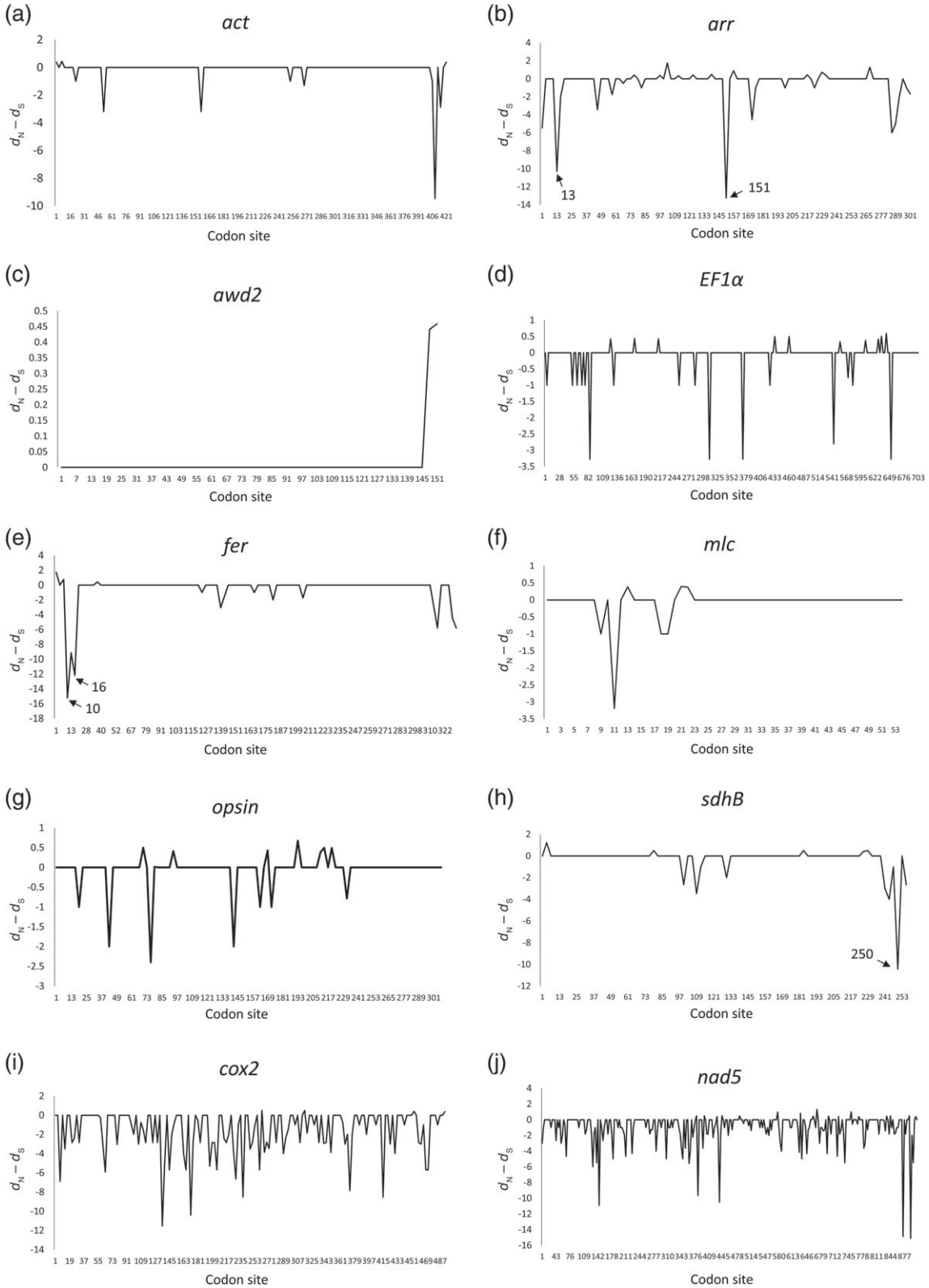

Fig. 2. Sliding window analyses of nonsynonymous–synonymous substitution ($d_N - d_S$) of nuclear (a–h) and representative mitochondrial (i, j) genes in *Euphaea*. The arrows indicate codon sites with $d_N - d_S < -10$.

Table 2. Summary of the parameter estimates of phylogenetic analyses of mitochondrial (Mt) and nuclear (Nu) genes in *Euphaea* species

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 noncoding									
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 uninformative sites.

PI sites, parsimony-informative sites; CI, consistency index; *Pi*, proportion of invariant sites; *α*, shape of gamma distribution; TS, transition; TV, transversion. The substitution rate categories refer to Fig. 4 and Appendix S7.

those of mitochondrial genes (protein-coding and noncoding) in the third codon sites ($t = 8.001$, $P < 0.001$). As expected, the third codons of mitochondrial protein-coding genes had the highest substitution rates. For noncoding mitochondrial genes, the substitution rate of the *A + T-rich* region was higher than that of *l-rRNA* and *s-rRNA*. The lowest substitution rate of the third codons of mitochondrial *nad4l* (2.0612) was still three times higher than the highest rate in the third codons of nuclear *fer* (0.663). The highest rate of the third codons of mitochondrial *atp6* (4.3913) was approximately 44 times higher than the lowest rate in the third codons of nuclear *awd2* (0.0999). We observed that the *A + T-rich* region had an intermediate substitution rate between those of the third codons of mitochondrial protein-coding genes and nuclear genes.

Rate heterogeneity among sites

Higher values of *α* (shape of gamma distribution) correspond to genes with less rate heterogeneity among sites (e.g. a more even distribution of rates among sites). Mitochondrial genes had higher *α* (by approximately two-fold) than the nuclear genes ($t = 3.814$, $P < 0.001$; Table 2; Fig. 3b). The *A + T-rich* region had the highest *α* (2.7769), corresponding to the gene with the most even distribution of rates among sites. Among mitochondrial protein-coding genes, *nad2* had the highest *α* (2.0078), which was about three times higher than the lowest *α*, found for *cox2* (0.7011). Nuclear introns and exons had a similarly low *α*, with the exons of *EF1α* having the lowest (0.3485). These results revealed that at the species level, mitochondrial genes have less

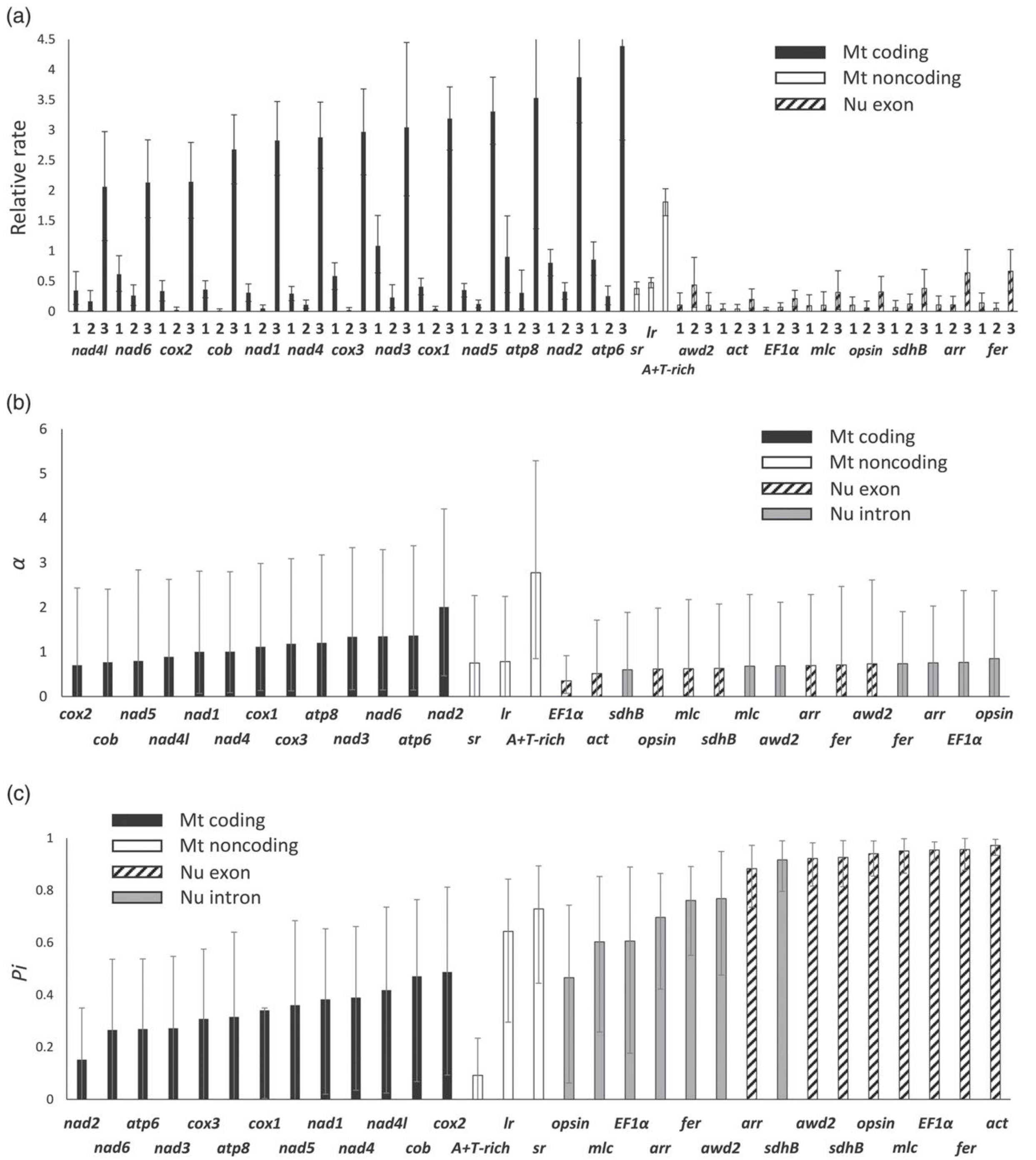

Fig. 3. Ranking of parameter estimates of phylogenetic analyses of mitochondrial protein-coding (Mt coding), noncoding (Mt noncoding) and nuclear (Nu) genes of *Euphaea* in MRBAYES. (a) Relative rates among codon sites estimated using the GTR + SSR model. (b) Shape of gamma distribution (α) estimated in the GTR + I + G model. (c) Proportion of invariant sites (*Pi*) estimated in the GTR + I + G model.

rate heterogeneity (more even rates) among sites than those of nuclear exons and introns.

Proportion of invariant sites

Among mitochondrial genes, *sr-RNA* had the highest *Pi* (0.7290), which was about five times higher than the lowest *Pi*, found for *nad2* (0.1520) (Table 2, Fig. 3c). For nuclear genes, the highest *Pi* of *act* (0.9724) was twice that of *opsin* introns (0.4658). The average *Pi* values of both nuclear introns (0.6884 ± 0.1459) and exons (0.9383 ± 0.0276) were higher than for all mitochondrial genes (0.3687 ± 0.1620) ($t = 7.787$, $P < 0.001$). Except for *arr*, the *Pi* of nuclear exons all

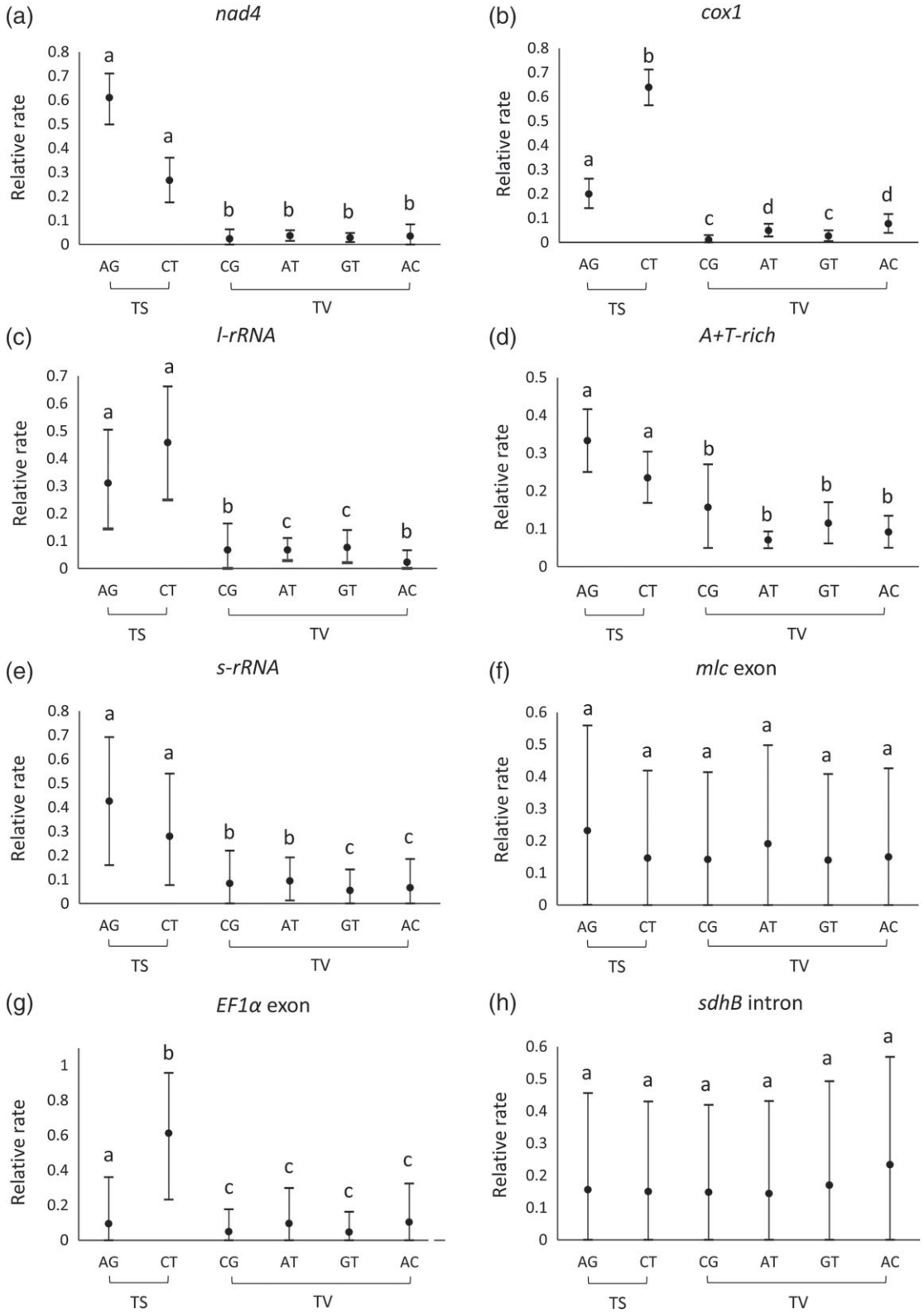

Fig. 4. The transformation rate matrices in selected mitochondrial and nuclear genes of *Euphaea* (mean ± 95% credibility interval, CI). The letters above the bars represent the rate categories estimated in the model selection of MEGA. TS, transition; TV, transversion.

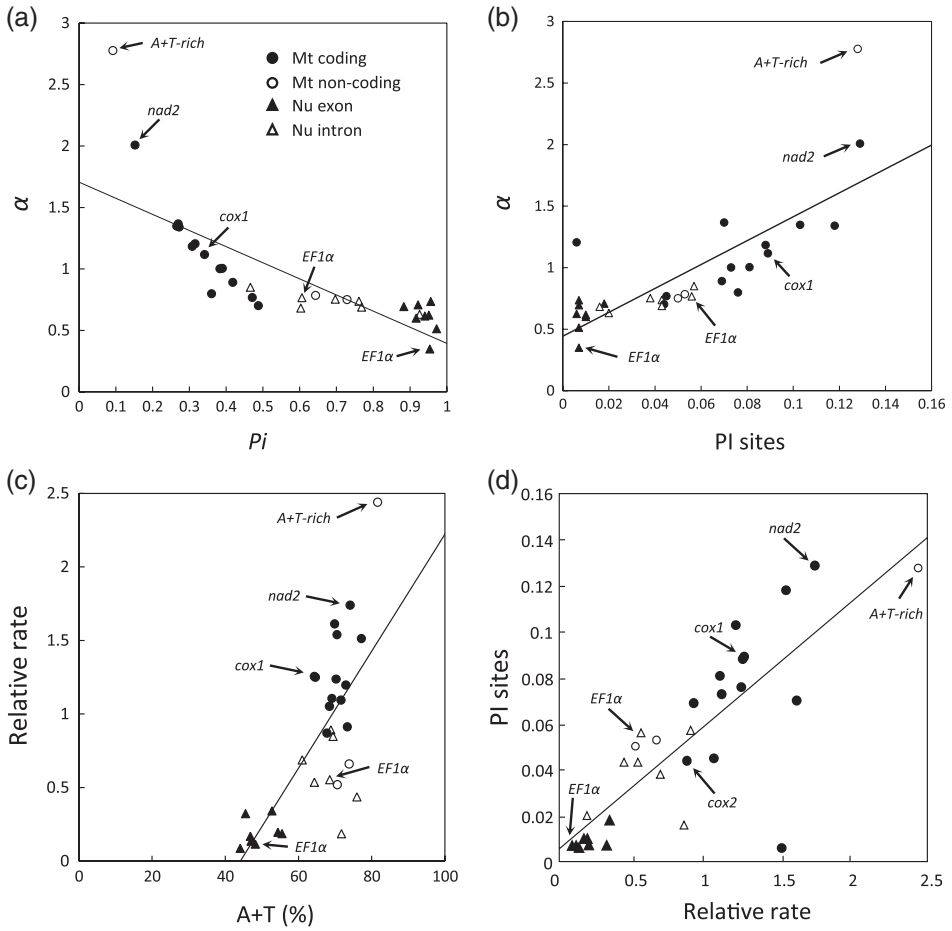

Fig. 5. Correlation analysis of phylogenetic parameter estimates of mitochondrial (Mt) and nuclear (Nu) genes in *Euphaea*. (a) Relationship between the shape of gamma distribution (α) and the estimated proportion of invariant sites (*Pi*). (b) Relationship between α and parsimony-informative (PI) sites. (c) The relationship between relative substitution rate and A + T bias. (d) Relationship between PI and relative substitution rate. All regressions are significant ($P < 0.001$).

approached 1 and were larger than the *Pi* of introns, suggesting that all sites in exons were nearly invariable at the species level. The *A + T-rich* region had the lowest *Pi* (0.0920) overall, indicating that approximately 90% of its sites were variable. The highest *Pi* of mitochondrial coding genes was observed for *cox2* (0.4877), which was three times higher than the lowest *Pi*, found in *nad2* (0.1520).

*Transformation rate matrices (*Q *matrix)*

Higher transition rates than transversion rates were observed for all genes analysed (Fig. 4, Appendix S7). Among mitochondrial genes, transition rates fell into different rate categories (AG vs CT) in *cox1*, *cox2*, *cox3* and *nad1*, suggesting that these genes were more asymmetric in transition rate distribution than the other genes, which had only one rate category (*nad2*, *atp8*, *atp6*, *nad3–6*, *nad4l*, *cob*, *l-rRNA*, *s-rRNA* and the *A + T-rich* region), whereas nuclear genes all had more symmetric transition rates (only one rate category), except for *EF1α* exons (Table 2, Fig. 4; Appendix S7). Among the transversion rates, mitochondrial *nad2*, *cox1*, *cox3*, *l-rRNA* and *s-rRNA* were more asymmetric (two rate categories), with the remaining mitochondrial genes having only one rate. In contrast, all nuclear exons and introns were inferred to have only one transversion rate category and were frequently in the same rate category as the transition rates (*arr* exons, *awd2* introns, *EF1α* introns, *mlc* exons, *opsin* exons, *sdhB* exons and *sdhB* introns), indicating that they had more symmetric rate matrices. The highest skew in rate matrix occurred in *cox1* and *cox3* (four rate categories, relative rates of CT = 0.6394 and 0.5701, respectively; Fig. 4; Appendix S7), which were nearly 63 and 30 times higher than the lowest rates of CG = 0.0101 and 0.0189, respectively, probably due to an excess of CT transitions.

Correlation among parameters

There was a significant negative correlation between α and *Pi* ($r^2 = 0.6017$, $P < 0.001$; Fig. 5a), suggesting that for genes with

more invariant sites (higher *Pi*), the remaining sites tended to show more rate heterogeneity (lower α). The *A + T-rich* region and *nad2* showed fewer invariant sites (lower *Pi*) and less rate heterogeneity (higher α) than the popular *cox1* and *Ef1α* introns. Values of α were significantly positively correlated with PI sites ($r^2 = 0.6112$, $P < 0.001$), indicating that genes with less rate heterogeneity (higher α) such as the mitochondrial *A + T-rich* region and *nad2* had more informative sites than many mitochondrial and nuclear genes (Fig. 5b). Relative substitution rates exhibited a significant positive correlation with A + T composition ($r^2 = 0.4944$; $p < 0.001$; Fig. 5c). PI sites were positively correlated with relative substitution rates ($r^2 = 0.6721$, $P < 0.001$) (Fig. 5d), indicating that the genes with higher relative substitution rates (the *A + T-rich* region and *nad2*) had more informative sites.

Discussion

The analysis of the four complete mitogenomes of *Euphaea* species provides useful information to guide the choice of mitochondrial markers for further phylogenetic and phylogeographic studies in damselflies. Our results indicate that mitochondrial protein-coding *nad2* and noncoding *A + T-rich* (control region) have the highest mutation rates and higher phylogenetic utility (higher PI, more parsimony-informative sites; higher α, less rate heterogeneity among sites; higher relative substitution rate) than all other mitochondrial and nuclear genes analysed. It is significant that *nad2* (990 bp) and the *A + T-rich* region (1109 bp) have a higher number of PI sites, given that many mitochondrial genes are longer (e.g. *nad5*, 1723 bp; *cox1*, 1548 bp). The *nad2* and the *A + T-rich* region should be selected to estimate species phylogeny and population genealogy in damselflies, given that financial and technological limitations constrain the use of multiple genes, longer sequences or whole-genome sequencing. These two mitochondrial genes had similar high values for CI (less homoplasy) compared with nuclear genes; however, the *nad2* and the *A + T-rich* region show slight disadvantages in their *Q* matrices (more asymmetry of rate categories) compared with the nuclear genes.

The mitochondrial gene *cox1* is now a widely used animal DNA barcoding marker for species identification (over 163 000 *cox1* sequences, release 6.50 – v1, 31 December 2015, the Barcode of Life Data System, http://www.boldsystems.org). However, the successful identification rates of using *cox1* for many related insect taxa were low (e.g. Meier *et al.*, 2006; Elias *et al.* 2007; reviewed in Jinbo *et al.* 2011), and the sequences of *cox1* were found to be under positive selection or selection relaxation [e.g. mirid bugs (Miridae); Wang *et al.*, 2017]. Our findings suggested that *cox1* was approximately average in all parameter estimates of phylogenetic performance and was sometimes outperformed by other mitochondrial genes (e.g. *nad3* and *nad6*) at the species and population levels. This result is surprising given that *nd3* and *nd6* are among the smallest mitochondrial genes, and *cox1* is among the largest. McDonagh *et al.* (2016) came to a similar conclusion that *cox1* might not be the most appropriate genetic marker based on a mitogenomic analysis of higher relationships of insects at the order level. The *cox1* and other mitochondrial genes (such as *12S*, *16S* and *cob*) that were used earlier in the field were not chosen because they are particularly informative phylogenetically, but rather, they were chosen because of the presence of highly conserved primers.

Although not as frequently used as the protein-coding *nad2* (e.g. Chen *et al.* 2017), the noncoding *A + T-rich* region has been shown to be a particularly suitable molecular marker in species-level phylogenetics and phylogeographic studies and in delimiting cryptic species, e.g. click beetles (Elateridae, Amaral *et al.*, 2017), butterflies (*Erebia*, Vila & Björklund 2004) and flies (*Drosophila*, Oliveira *et al.* 2007). Our results confirmed that the *A + T-rich* region exhibits the best quality for phylogeographic and population analyses, including the highest α, PI sites, and relative substitution rate for resolving recently diverged species and populations. Currently the *A + T-rich* region is rarely used in phylogeographic and population genetic studies of damselflies (Ballare & Ware 2011; Bybee *et al.* 2016), probably due to the unavailability of taxon-specific primers, the difficulty of directly sequencing the *A + T-rich* region, which has numerous tandem repeats, and the analytical problems of highly variable lengths and indels. Nevertheless, the advancement of next-generation sequencing technologies (Goodwin *et al.* 2016), the collection of conserved primers and amplification protocols (reviewed in Simon *et al.* 1994, 2006) and the accelerated accumulation of complete damselfly mitochondrial genomes (e.g. Lin *et al.* 2010; Lorenzo-Carballa *et al.* 2014, 2016; Chen *et al.* 2015; Wang *et al.* 2015; Feindt *et al.* 2016a, b) will greatly facilitate the development of taxon-specific *nad2* and *A + T-rich* region primers in diverse damselfly species. Determining the phylogenetic utility of these two mitochondrial markers in other damselflies requires further studies.

Apart from the frequently used nuclear loci such as ribosomal genes (*18S*, *5.8S*, *28S* and *ITS*), histone 3 (H3) and *EF1α* (Ballare & Ware 2011), few new nuclear genetic markers for damselflies have been developed for phylogeography and population genetic studies except for the exon-primed intron-crossing (EPIC) markers in *Euphaea* (Lee & Lin 2012a) and *Coenagrion* (Ferreira *et al.* 2014). The development of species-specific EPIC primers or microsatellites containing single nucleotide polymorphisms (SNPs) is time-consuming and may not be easily applied across a range of taxa. Until now, phylogeographic and population genetic studies of damselflies have remained impeded by the lack of a set of high-resolution genetic markers that can be easily applied across species. Mitochondrial *nad2* and the *A + T-rich* region therefore provide efficient options as high-resolution markers for the first empirical screen of genetic diversity in targeted damselfly species and populations.

Our findings indicated that most mitochondrial and nuclear genes in *Euphaea* damselflies experienced frequent purifying selection, except for two cases of potential positive selection in *EF1α* of *E. ornata* and *nad3* at the species level. However, the putative signatures of positive selection identified in *EF1α* and *nad3* should be treated with caution because of the limitation of our sample sizes. The main selective force driving the sequence variation of most mitochondrial genes in *Euphaea* was

purifying selection rather than positive selection (genetic draft; Bazin *et al.*, 2006, Mulligan *et al.*, 2006). This is not surprising given that the proteins transcribed by mitochondrial genes are used in electron transport chains that are important in energy metabolism. Therefore, most mutations in mitochondrial genes are probably deleterious and removed by strong purifying selection (Boore 1999; Ballard & Whitlock 2004). We found evidence to show that mitochondrial genes had experienced stronger purifying selection than nuclear genes in *Euphaea* species, given that the effective population size of mitochondrial genes is only a quarter of that of nuclear genes. However, the estimated strength of selection in mitochondrial genes did not correlate with the population sizes of the *Euphaea* species, suggesting that factors other than population size (neutral evolution), such as extreme purifying selection, lineage-specific adaptive selection and life history, may work together to shape mitochondrial genetic diversity (Stoeckle & Thaler 2014; Ellegren & Galtier 2016).

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1. *Euphaea* specimens and GenBank accession numbers of mitochondrial and nuclear genes used in this study.

Appendix S2. Distributional map and species phylogeny of the four *Euphaea* species reconstructed based on 13 protein-coding, two ribosomal, and the *A + T-rich* genes of mitochondrial genomes, and eight nuclear genes in BEAST.

Appendix S3. Pairwise sequence divergence of mitochondrial and nuclear genes of *Euphaea* calculated using the Tajima–Nei model in MEGA.

Appendix S4. Mutation rates of mitochondrial genes of *Euphaea* calculated using the Tajima–Nei model in MEGA.

Appendix S5. The d_N/d_S ratios (ω) of mitochondrial and nuclear genes of *Euphaea* calculated using the branch model in the codeml option of PAML.

Appendix S6. The d_N and d_S ratios of the sequences with species-specific ω in *Euphaea*.

Appendix S7. The transformation rate matrices of mitochondrial and nuclear genes of *Euphaea*.

Acknowledgements

We are grateful to Chris Simon and one anonymous reviewer for their comments and suggestions. This research project was supported by the National Science Council (NSC 100-2311-B-029-004-MY3) and Ministry of Science and Technology (MOST 104-2621-B-003 -002 -MY3) of Taiwan. CPL thanks Kyoto University for a visiting professorship during the writing of the manuscript.

Author contributions

CPL designed the study; MYC, JFW and APL collected the data; YCC, MYC and JFW analysed data; and YCC and CPL wrote the manuscript.

References

Amaral, D.T., Mitani, Y., Oliveira, G., Ohmiya, Y. & Viviani, V.R. (2017) Revisiting Coleoptera A+ T-rich region: structural conservation, phylogenetic and phylogeographic approaches in mitochondrial control region of bioluminescent Elateridae species (Coleoptera). *Mitochondrial DNA Part A*, **28**, 671–680.

Avise, J.C. (2004) *Molecular Markers, Natural History, and Evolution*. Sinauer Associates, Sunderland, Massachusetts.

Ballard, J.W.O. & Whitlock, M.C. (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729–744.

Ballare, E.F. & Ware, J.L. (2011) Dragons fly, biologists classify: an overview of molecular odonate studies, and our evolutionary understanding of dragonfly and damselfly (Insecta: Odonata) behavior. *International Journal of Odonatology*, **14**, 137–147.

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

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

Bybee, S., Córdoba-Aguilar, A., Duryea, M.C. *et al.* (2016) Odonata (dragonflies and damselflies) as a bridge between ecology and evolutionary genomics. *Frontiers in Zoology*, **13**, 46.

Cameron, S.L. (2014) Insect mitochondrial genomics: implications for evolution and phylogeny. *Annual Review of Entomology*, **59**, 95–117.

Caravas, J. & Friedrich, M. (2013) Shaking the Diptera tree of life: performance analysis of nuclear and mitochondrial sequence data partitions. *Systematic Entomology*, **38**, 93–103.

Chen, M.Y., Chaw, S.M., Wang, J.F., Villanueva, R.J.T., Nuneza, O.M. & Lin, C.P. (2015) Mitochondrial genome of a flashwing demoiselle, *Vestalis melania* from the Philippine archipelago. *Mitochondrial DNA*, **26**, 720–721.

Chen, Y.T., Tseng, H.Y., Jeng, M.L., SU, Y.C., Huang, W.S. & Lin, C.P. (2017) Integrated species delimitation and conservation implications of an endangered weevil *Pachyrhynchus sonani* (Coleoptera: Curculionidae) in green and Orchid Islands of Taiwan. *Systematic Entomology*, **42**, 796–813.

Chesters, D. (2017) Construction of a species-level tree of life for the insects and utility in taxonomic profiling. *Systematic Biology*, **66**, 426–439.

Drummond, A.J., Suchard, M.A., Xie, D. & Rambaut, A. (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.

Elias, M., Hill, R.I., Willmott, K.R., Dasmahapatra, K.K., Brower, A.V.Z., Mallet, J. & Jiggins, C.D. (2007) Limited performance of DNA barcoding in a diverse community of tropical butterflies. *Proceedings of the Royal Society of London B: Biological Sciences*, **274**, 2881–2889.

Ellegren, H. & Galtier, N. (2016) Determinants of genetic diversity. *Nature Reviews Genetics*, **17**, 422–433.

Feindt, W., Herzog, R., Osigus, H.-J., Schierwater, B. & Hadrys, H. (2016a) Short read sequencing assembly revealed the complete

mitochondrial genome of *Ischnura elegans* Vander Linden, 1820 (Odonata: Zygoptera). *Mitochondrial DNA Part B*, **1**, 574–576.

Feindt, W., Osigus, H.-J., Herzog, R., Mason, C.E. & Hadrys, H. (2016b) The complete mitochondrial genome of the neotropical helicopter damselfly *Megaloprepus caerulatus* (Odonata: Zygoptera) assembled from next generation sequencing data. *Mitochondrial DNA Part B*, **1**, 497–499.

Ferreira, S., Lorenzo-Carballa, M.O., Torres-Cambas, Y., Cordero-Rivera, A., Thompson, D.J. & Watts, P.C. (2014) New EPIC nuclear DNA sequence markers to improve the resolution of phylogeographic studies of coenagrionids and other odonates. *International Journal of Odonatology*, **17**, 135–147.

Frankham, R. (1996) Relationship of genetic variation to population size in wildlife. *Conservation Biology*, **10**, 1500–1508.

Funk, D.J. & Omland, K.E. (2003) Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution and Systematics*, **34**, 397–423.

Galtier, N., Nabholz, B., Glémin, S. & Hurst, G.D.O. (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, **18**, 4541–4550.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D. & Bairoch, A. (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, **31**, 3784–3788.

Gissi, C., Iannelli, F. & Pesole, G. (2008) Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity*, **101**, 301–320.

Goodwin, S., McPherson, J.D. & McCombie, W.R. (2016) Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, **17**, 333–351.

Grimaldi, D. & Engel, M.S. (2005) *The Evolution of the Insects*. Cambridge University Press, New York, New York.

Hebert, P.D.N., Cywinska, A., Ball, S.L. & deWaard, J.R. (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B*, **270**, 313–321.

Huang, J.P. & Lin, C.P. (2011) Lineage-specific late pleistocene expansion of an endemic subtropical gossamer-wing damselfly, *Euphaea formosa*, in Taiwan. *BMC Evolutionary Biology*, **11**, 94.

Huelsenbeck, J.P. & Ronquist, F.R. (2001) MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**, 754–755.

Jinbo, U., Kato, T. & Ito, M. (2011) Current progress in DNA barcoding and future implications for entomology. *Entomological Science*, **14**, 107–124.

Lee, Y.H. & Lin, C.P. (2012a) Pleistocene speciation with and without gene flow in *Euphaea* damselflies of subtropical and tropical east Asian islands. *Molecular Ecology*, **21**, 3739–3756.

Lee, Y.H. & Lin, C.P. (2012b) Morphometric and genetic differentiation of two sibling gossamer–wing damselflies, *Euphaea formosa* and *E. yayeyamana*, and adaptive trait divergence in subtropical east Asian islands. *Journal of Insect Science*, **12**, 53.

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

Lin, C.P. & Danforth, B.N. (2004) How do insect nuclear and mitochondrial genes substitution patterns differ? Insights from Bayesian analysis of combined datasets. *Molecular Phylogenetics and Evolution*, **30**, 686–702.

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

López-López, A. & Vogler, A.P. (2017) The mitogenome phylogeny of Adephaga (Coleoptera). *Molecular Phylogenetics and Evolution*, **114**, 166–174.

Lorenzo-Carballa, M.O., Thompson, D.J., Cordero-Rivera, A. & Watts, P.C. (2014) Next generation sequencing yields the complete mitochondrial genome of the scarce blue-tailed damselfly, *Ischnura pumilio*. *Mitochondrial DNA*, **25**, 247–248.

Lorenzo-Carballa, M.O., Tsubaki, Y., Plaistow, S.J. & Watts, P.C. (2016) The complete mitochondrial genome of the broad-winged damselfly *Mnais costalis* Selys (Odonata: Calopterygidae) obtained by next-generation sequencing. *International Journal of Odonatology*, **19**, 191–198.

McDonagh, L.M., West, H., Harrison, J.W. & Stevens, J.R. (2016) Which mitochondrial gene (if any) is best for insect phylogenetics? *Insect Systematics & Evolution*, **47**, 245–266.

Meier, R., Shiyang, K., Vaidya, G. & NG, P.K.L. (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology*, **55**, 715–728.

Meiklejohn, C.D., Montooth, K.L. & Rand, D.M. (2007) Positive and negative selection on the mitochondrial genome. *Trends in Genetics*, **23**, 259–263.

Miraldo, A., Li, S., Borregaard, M.K. *et al.* (2016) An Anthropocene map of genetic diversity. *Science*, **353**, 1532–1535.

Misof, B., Liu, S., Meusemann, K. *et al.* (2014) Phylogenomics resolves the timing and pattern of insect evolution. *Science*, **346**, 763–767.

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

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

Neher, R.A. (2013) Genetic draft, selective interference, and population genetics of rapid adaptation. *Annual Review of Ecology, Evolution, and Systematics*, **44**, 195–215.

Oliveira, M.T., Azeredo-Espin, A.M. & Lessinger, A.C. (2007) The mitochondrial DNA control region of Muscidae flies: evolution and structural conservation in a dipteran context. *Journal of Molecular Evolution*, **64**, 519–527.

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

Piganeau, G. & Eyre-Walker, A. (2009) Evidence for variation in the effective population size of animal mitochondrial DNA. *PLoS One*, **4**, e4396.

Rambaut A, Suchard M, Xie D, Drummond A (2014) Tracer v1.6. http://beast.bio.ed. ac.uk/Tracer.

Roux, J., Privman, E., Moretti, S., Daub, J.T., Robinson-Rechavi, M. & Keller, L. (2014) Patterns of positive selection in seven ant genomes. *Molecular Biology and Evolution*, **31**, 1661–1685.

Rubinoff, D. & Holland, B.S. (2005) Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Systematic Biology*, **54**, 952–961.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. & Flook, P. (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, **87**, 651–701.

Simon, C., Buckley, T.R., Frati, F., Stewart, J.B. & Beckenbach, A.T. (2006) Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. *Annual Review of Ecology, Evolution and Systematics*, **37**, 545–579.

Stoeckle, M.Y. & Thaler, D.S. (2014) DNA barcoding works in practice but not in (neutral) theory. *PLoS One*, **9**, e100755.

Sunnucks, P. (2000) Efficient genetic markers for population biology. *Trends in Ecology and Evolution*, **15**, 199–203.

Tajima, F. & Nei, M. (1984) Estimation of evolutionary distance between nucleotide sequences. *Molecular Biology and Evolution*, **1**, 269–285.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, **30**, 2725–2729.

Vila, M. & Björklund, M. (2004) The utility of the neglected mitochondrial control region for evolutionary studies in Lepidoptera (Insecta). *Journal of Molecular Evolution*, **58**, 280–290.

Wang, J.F., Chen, M.Y., Chaw, S.M., Morii, Y., Yoshimura, M., Sota, T. & Lin, C.P. (2015) Complete mitochondrial genome of an enigmatic dragonfly, *Epiophlebia superstes* (Odonata, Epiophlebiidae). *Mitochondrial DNA*, **26**, 718–719.

Yamauchi, M.M., Miya, M.U. & Nishida, M. (2004) Use of a PCR-based approach for sequencing whole mitochondrial genomes of insects: two example (cockroach and dragonfly) based on the method developed for decapod crustaceans. *Insect Molecular Biology*, **13**, 435–442.

Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**, 1586–1591.

Accepted 1 March 2018
First published online 11 April 2018