

Pleistocene speciation with and without gene flow in *Euphaea* damselflies of subtropical and tropical East Asian islands

YAT-HUNG LEE and CHUNG-PING LIN

Department of Life Science & Center for Tropical Ecology and Biodiversity, Tunghai University, Taichung 40704, Taiwan

Abstract

Climatic oscillations during the Pleistocene period could have had a profound impact on the origin of tropical species by the alternation of allopatric isolation and interpopulation gene flow cycles. However, whether tropical speciation involves strictly allopatric isolation, or proceeds in the face of homogenizing gene flow, is relatively unclear. Here, we investigated geographical modes of speciation in four closely related *Euphaea* damselfly species endemic to the subtropical and tropical East Asian islands using coalescent analyses of a multilocus data set. The reconstructed phylogenies demonstrated distinct species status for each of the four species and the existence of two sister species pairs, *Euphaea formosa*/*E. yayeyamana* and *E. decorata*/*E. ornata*. The species divergence time of the sibling *Euphaea* damselflies dates back to within the last one Mya of the Middle to Lower Pleistocene. The speciation between the populous *E. formosa* of Taiwan and the less numerous *E. yayeyamana* of the Yaeyama islands occurred despite significant bidirectional, asymmetric gene flow, which is strongly inconsistent with a strictly allopatric model. In contrast, speciation of the approximately equal-sized populations of *E. decorata* of the southeast Asian mainland and *E. ornata* of Hainan is inferred to have involved allopatric divergence without gene flow. Our findings suggest that differential selection of natural or sexual environments is a prominent driver of species divergence in subtropical *E. formosa* and *E. yayeyamana*; whereas for tropical *E. decorata* and *E. ornata* at lower latitudes, allopatric isolation may well be a pivotal promoter of species formation.

Keywords: allopatric isolation, divergence with migration, Euphaeidae, gene flow, geography of speciation, Odonata

Received 3 January 2012; revision received 5 April 2012; accepted 18 April 2012

Introduction

Climatic oscillations during the Pleistocene ice ages had profound impacts on the origin and persistence of extant species (Hewitt 2000). These severe glacial periods produced enormous changes in species distributions, which led to extinction over large parts of their range, dispersion and colonization into new habitats and survival and subsequent expansion of isolated remnant populations (Hewitt 2000). The Pleistocene glacial cycles had large effects on the origin of tropical species,

dependent largely on the latitude, topography and life history of the species in question (Hewitt 2000). For example, in the tropical lowland forests, the speciation process would have been determined predominantly by the alternation of allopatric isolation and interpopulation gene flow cycles caused by recurrent glacial contractions and interglacial expansions of forest patches (Moritz *et al.* 2009; Rull 2011). As for tropical high mountains, repeated altitudinal migrations during the glacial cycles could have promoted speciation by providing opportunities for successive allopatric isolation, secondary contact and consequent gene flow among populations (Rull 2011). However, studies have suggested that the majority of tropical speciation events

Correspondence: Chung-Ping Lin, Fax: 886 4 2359 0296; E-mail: treehops@thu.edu.tw

predate the Pleistocene in the pre-Quaternary (Rull 2008; Hoorn *et al.* 2010) and that extensive gene flow of mixing populations during the Pleistocene interglacial periods would have hindered the speciation process (e.g. Knowles 2000). This questions the extent to which the Pleistocene glacial cycle affected species formation in the tropics. Did tropical speciation involve mainly allopatric isolation without gene exchange or did it proceed in the face of recurrent homogenizing gene flow resulting from glacial cycles?

The role of gene flow in the process of species formation is contentious among evolutionary biologists (Coyne & Orr 2004; Hey 2006; Nosil 2008). Allopatric speciation remains the predominant speciation hypothesis in natural populations (Mayr 1954, 1963; reviewed in Coyne & Orr 2004). In the allopatric model of speciation, physical barriers fully disrupt gene flow between isolated populations (immigration rate, $m = 0$) until the completion of reproductive isolation as a by-product of genetic divergence caused by local adaptation, mutation and genetic drift (Gavrilets 2004). Other geographic models of speciation that permit selection, mutation or genetic drift-driven divergence in the presence of gene flow, such as parapatric ($0 < m < 0.5$) and sympatric ($m = 0.5$) speciation (Gavrilets 2003, 2004), have been perceived as playing minor roles because of the homogenizing effects of gene flow in diverged populations (Coyne & Orr 2004). The allopatric mode of speciation was based on the 'biological species concept', which in essence considered the whole genome of an individual as the unit of speciation (Mayr 1954, 1963). In contrast, Wu (2001) proposed the 'porous model' of speciation, which allows a portion of neutral genes within the genome to exchange continually between diverging populations, but genes that are differentially adapted between populations (i.e. speciation genes) are not exchangeable owing to a detrimental effect on reproduction (Wu 2001). Under this genic view of speciation, strongly divergent selection of a small number of genes from different environments may drive speciation in the presence of homogenizing gene flow (Hey 2006; Nosil 2008). Hey & Wakeley (1997) proposed the isolation model of speciation, in which fixed genetic differences accumulate and shared genetic variations are reduced by genetic drift through time during the speciation process. This model can be used as a null hypothesis to assess the validity of 'strictly' allopatric speciation in the complete absence of gene flow. Nielsen & Wakeley (2001) modified this isolation model into the isolation with migration (IM) model, which allows reciprocal gene exchanges to occur between diverging species. Using the IM model, the coalescence of multiple loci can be modelled for all possible genealogies between the diverging populations to estimate demographic parame-

ters, including effective population size, migration rate and divergence time and to test whether or not closely related species have exchanged genes during species formation (Hey 2006; Hey & Nielsen 2007).

Euphaea decorata, *E. formosa*, *E. ornata* and *E. yayeyamana* are four closely related damselflies (Insecta, Odonata and Euphaeidae) characterized by males having transparent forewings and black bands on their hind wings. The current distributions of the four *Euphaea* species are entirely allopatric (Fig. 1). *Euphaea formosa* and *E. yayeyamana* are endemic to subtropical Taiwan and the Yaeyama islands of the Ryukyu Archipelago, respectively (Matsuki & Lien 1978, 1984; Hayashi 1990) (Fig. 1). *Euphaea decorata* is widely distributed on the southeast Asian mainland, whereas *E. ornata* is only found on the tropical Hainan island (Dudgeon 1989). Oceanic barriers have been considered an effective isolation mechanism for these medium-sized *Euphaea* damselflies, which have short, rounded wings and limited long-distance dispersal (Dudgeon 1989; Corbet 1999). Allopatric species formation is proposed to have happened for these insular endemics, resulting from geographic isolation between the islands and the nearby Asian continent (Ota 1998; Voris 2000). The evidence for allopatric species has been based largely on fauna similarity among islands and the existence of parapatric or peripatric distributions of closely related species (Ota 1998). However, allopatric modes of speciation inferred from current species ranges can be misleading in that the contemporary range of a species often does not reflect its historical distribution at the time of speciation, and the geographical barriers may not have been

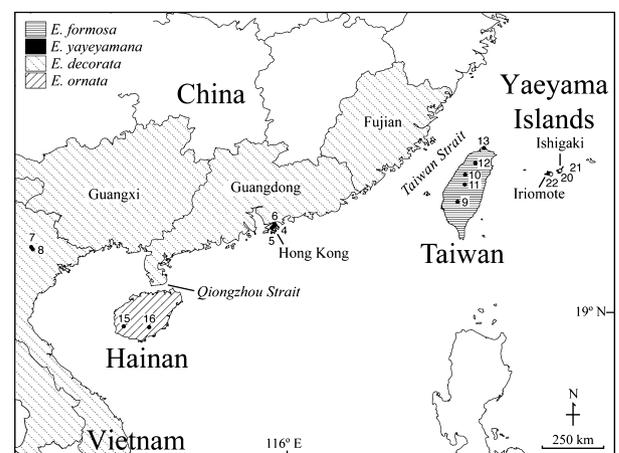


Fig. 1 A map of East Asia showing the currently known geographical distribution and sampling sites (in numbers) of the four *Euphaea* damselflies. The numbers correspond to the collecting information listed in Table S1 (Supporting information).

consistently present throughout the species history (Loros & Glor 2003).

The continental islands of southeast Asia have experienced recurrent sea level changes during past glacial cycles (Voris 2000). As sea levels fell during glacial periods, land bridges were formed, periodically connecting the Ryukyu Archipelago, Taiwan, Hainan and the nearby Asian continent (Voris 2000) and thus allowed intermittent secondary contacts among formerly isolated populations. Consequently, there could have been plenty of opportunities for resumed gene flow between these diverged *Euphaea* species throughout the glacial cycles. All four *Euphaea* species appear to have similar ecological characteristics, preferring fast-flowing forest streams with open canopy (Dudgeon 1989; Hayashi 1990; Huang & Lin 2011). Nevertheless, *E. formosa* and *E. yayeyamana* differ substantially in availability of larval prey, body size (Hayashi 1990) and wing shape (Lee & Lin 2012), which are linked to energetic costs and manoeuvre ability in flight. The decreased body size, long forewings and narrow hind wings of *E. yayeyamana* may represent an insular adaptation to limited larval prey and a reduced territorial competition in the smaller Yaeyama islands (Lee & Lin 2012). Local adaptation to a lower quality habitat is therefore probably to have driven the divergence of *E. yayeyamana* from its sister species, *E. formosa* (Hayashi 1990; Lee & Lin 2012). These insular *Euphaea* species therefore provide the opportunity to test the likelihood of assumed examples of allopatric speciation, by examining the dynamics of gene flow between species that are probably to have experienced glacial expansions and consequent episodic secondary contact, but have ample evidence to suggest that local adaptation could have played a major role in driving divergence.

Theoretical models have demonstrated that speciation in the face of gene flow is feasible under particular circumstances (Gavrilets 2003, 2004). Recently, empirical evidence for speciation with gene flow has been rapidly accumulating thanks to analytical advances (Hey 2006; Hey & Nielsen 2007) and the availability of a mass of genomic data (Ekblom & Galindo 2010). Significant signals of gene flow during species formation have been documented in a number of vertebrates, including chimpanzees (Osada & Wu 2005; Won & Hey 2005), gorillas (Osada & Wu 2005; Thalmann *et al.* 2007), lizards (Leaché & Mulcahy 2007), salamanders (Niemiller *et al.* 2008; Nadachowska & Babik 2009), fishes (Berner *et al.* 2009) and birds (Peters *et al.* 2007; Brumfield *et al.* 2008; Li *et al.* 2010; Yeung *et al.* 2011). These divergent population genetic studies have demonstrated significant gene flow among diverging natural populations and species. Studies of insects such as grasshoppers

(Carstens & Knowles 2007), butterflies (Bull *et al.* 2006; Arias *et al.* 2008), mosquitoes (Turner *et al.* 2005) and fruit flies (Hey & Nielsen 2004) have also revealed substantial migration among diverging species that supports the model of speciation in the face of gene flow. Despite recent advances and emerging patterns of speciation in a few selected model organisms and a small number of tropical species, the vast majority of 'non-model' organisms on Earth inhabit tropical regions, where the prevailing mode of speciation remains largely unknown. To address fully the relative frequency and importance of different evolutionary processes leading to speciation, it is essential to understand the dynamics of gene flow between recently diverged species across taxa and regions, especially in nonmodel organisms in the tropics.

We present an investigation on the evolutionary histories and divergent population genetics of *E. decorata*, *E. formosa*, *E. ornata* and *E. yayeyamana* using multilocus sequences. This study is the first to test the hypotheses that *E. formosa* and *E. yayeyamana*, and *E. decorata* and *E. ornata* are more closely related to each other as two sibling species pairs than to the other East Asian *Euphaea* species and that these four damselflies represent genetically differentiated species among which gene flow ceased completely some time ago, after isolation. Second, we have examined the role of past climate change in speciation in subtropical and tropical East Asian islands by testing the prediction that the speciation in these four *Euphaea* species occurred during the Pleistocene period. Our final aim was to quantify the changes of effective population sizes and the level of gene flow during species formation between *E. formosa* and *E. yayeyamana*, and between *E. decorata* and *E. ornata* and to reveal the geographic mode of speciation by testing the null hypothesis of a strict allopatry model against a series of nested demographic models in IM.

Materials and methods

Sampling, DNA extraction and sequencing

Damselfly specimens of the four *Euphaea* species (*decorata*, *formosa*, *ornata* and *yayeyamana*) were collected from two to five geographic localities across the species' ranges (Fig. 1, Table S1, Supporting information). At least three individuals per species were chosen from each location, giving a total of 109 individuals for the four *Euphaea* species (25 *decorata*, 24 *formosa*, 29 *ornata* and 31 *yayeyamana*). To reconstruct the phylogenetic relationship of the East Asian *Euphaea* species, the specimens of six additional *Euphaea* species (*amphicyana*, *cora*, *masoni*, *querini*, *refulgens* and *subcostalis*) and *Bayadera*

brevicauda (Selys, 1853) were obtained from Indonesia, the Philippines, Taiwan and Vietnam (Table S1, Supporting information). *Bayadera brevicauda* was used as an outgroup in the phylogenetic analyses because the genus *Bayadera* is sister to *Euphaea* (Bybee *et al.* 2008). The collected specimens were immersed in 95% EtOH immediately after capture in the field and later stored at -80°C in the laboratory. Genomic DNA was extracted from the insect's thoracic muscle or hind femur using a MasterPure™ Complete DNA Purification Kit (EPICENTRE, Wisconsin, USA). Voucher specimens were deposited into the insect collection of the Department of Life Science at Tunghai University.

For all individuals of the four *Euphaea* species (*decorata*, *formosa*, *ornata* and *yayeyamana*), DNA fragments of the mitochondrial *cox2* gene were amplified using published primers (C2-J-3102, Jordan *et al.* 2003; E-C2-N-3740, Huang & Lin 2011). The *Euphaea*-specific ND5-143F (5'-CCA TGA TCA AAT CTC TTA ACT AA-3') and ND5-1133R (5'-TGC TGC TAT RAC YAA RAG TGC-3') primers were designed and used to amplify fragments of the mitochondrial *nad5* gene. A eukaryote-specific 18SF/28SR primer set was used to amplify fragments of the ribosomal *ITS* gene (Weekers *et al.* 2001). The fragments of nine additional nuclear loci (*act*, *anon*, *arr*, *awd2*, *EF1 α* , *fer*, *mlc*, *lop1* and *sdhB*) were amplified using EPIC (exon-primed, intron-crossing) primers developed for *Euphaea* by Lee & Lin (Molecular Ecology Resources Primer Development Consortium *et al.* 2011). For individuals of the other six *Euphaea* species (*amphicyana*, *cora*, *masoni*, *querini*, *refulgens* and *subcostalis*) and *B. brevicauda*, fragments of the *cox2* and *arr* genes were amplified. Each PCR contained a total volume of 50 μL , composed of 100–300 ng of template DNA, 0.4 μM of forward and reverse primers, 0.2 mM of dNTPs and 0.04 unit of ProTaq polymerase (Protech Technology, Taiwan). The PCR profile was as follows: denaturing at 94°C for 3 min, 35 cycles of amplification at 94°C for 1 min followed by the optimal annealing temperature (*cox2* & *nad5*: 54°C ; *ITS*: 52°C ; EPIC loci: Lee & Lin Molecular Ecology Resources Primer Development Consortium *et al.* 2011) for 1 min and 72°C for 1 min, then a final extension at 72°C for 10 min. The target PCR products were purified using a Gel/PCR DNA Fragments Extraction Kit (Geneaid, Taipei, Taiwan) and either sequenced directly (*act*, *awd2*, *cox2*, *EF1 α* , *ITS*, *nad5*, *sdhB* and *anon*), or cloned into the pCR® 2.1-TOPO vector (Invitrogen) and sequenced by randomly choosing more than three positive clones (*arr2*, *fer*, *mlc* and *lop1*). Positive clones were confirmed by performing PCRs using M13F/M13R primers. Purified PCR products and plasmid DNA were sequenced in both directions on an ABI PRISM™ 377 automatic sequencer (Perkin Elmer, USA) at the Mission Biotech, Taiwan.

The DNA sequences were assembled and manually edited in EditSeq (DNASTAR Lasergene package, Madison, USA). The edited sequences were aligned using the CLUSTAL W method in MegAlign (DNASTAR; Lasergene Package, Madison, USA) and then modified manually. The sequence alignments were unambiguous as they contained only short indels. Gametic phases of nuclear sequences with heterozygous sites were determined using a PHASE algorithm of a coalescent-based Bayesian method (Stephens *et al.* 2001) implemented in DNASP (v. 5, Librado & Rozas 2009). The Markov chain Monte Carlo (MCMC) process was set to 100 000 generations and sampled every 100 generations to infer haplotype data sets for subsequent analyses.

Genetic diversity statistics and phylogenetic analyses

Standard population genetic diversity statistics were calculated for each locus within each of the four *Euphaea* species (*decorata*, *formosa*, *ornata* and *yayeyamana*) using DNASP: the number of polymorphic sites (*S*), the haplotype diversity (H_d) and nucleotide diversity (π). The net genetic distance per site per locus between sibling *Euphaea* species was calculated using the Tamura–Nei substitution model (Tamura & Nei 1993) in MEGA 5 (Tamura *et al.* 2011) to account for heterogeneity in base frequencies and substitution rates among sites. The migration rate (Nm) of each locus was calculated using the method of Hudson *et al.* (1992) implemented in DNASP.

The phylogenetic analysis was performed on the combined *cox2* and *arr* sequences of representative *Euphaea* species (Table S1, Supporting information, Fig. 1) using Maximum Parsimony (MP) in PAUP* (v. 4.0b10, Swofford 2002), with a parsimony ratchet algorithm (Nixon 1999) implemented in PAUPRAT (v. 1, Sikes & Lewis 2001). Twenty independent ratchet analyses with the default setting of 200 replicates were run. Parsimony branch supports were calculated using bootstrapping of 1000 replicates of tree bisection and reconnection (TBR) branch swapping, with 100 replications of random sequence addition. For Maximum Likelihood (ML) and Bayesian Inference (BI), the best-fitting model of nucleotide substitution (*cox2*: TPM2uf+G; *arr*: HKY+G) and prior values of parameters were calculated in JMODELTEST (v. 0.1.1, Posada 2008) using the Akaike Information Criterion (AIC). ML tree searches and parameter optimization were conducted using a rapid approximation algorithm implemented in RAXML (Stamatakis 2006). The ML tree was reconstructed using the GTRGAMMA model to accommodate rate heterogeneity within each gene using four discrete rate categories of a gamma distribution (GAMMA). We conducted 1000 iterations in each ML analysis, and identified the optimal ML tree

by comparing the likelihood values among them. To assess the support for internal nodes of the ML tree, we calculated 1000 likelihood bootstrap (LB) replications. BI analyses were carried out using MRBAYES (v. 3.12, Huelsenbeck & Ronquist 2001). Two independent Bayesian runs, each with four Markov chains, were performed simultaneously. The MCMC chains were run for 2×10^8 generations, with trees being sampled every 2×10^4 generations. MCMC processes were terminated after the average split frequencies fell below 0.01, suggesting convergence of the separate runs. One quarter of the sampled trees (2500) were discarded as burn-in. The Bayesian posterior probability (BPP) was calculated by constructing a 50% majority consensus tree from the remaining trees in PAUP*.

A *Euphaea* species tree was also co-estimated from the joint posterior probability of *cox2* and *arr* gene trees using *BEAST (v. 1.6.1, Heled & Drummond 2010), with multiple individuals per species (Fig. S1, Supporting information). The substitution model, clock model and tree model of the two genes were set as unlinked. A random local clock and default Yule process of the species tree prior were used. Two independent MCMC runs were performed for 1×10^8 generations, with trees sampled every 1×10^4 generations, and the first 10% of the runs (1×10^7 generations) being discarded as burn-in. The convergence of runs was determined by the effective sample size (ESS) values of parameters (>100), likelihood scores through time plot using TRACER (v. 1.5, Drummond & Rambaut 2007), and by examining the posterior split probabilities and branch lengths in AWTY (Wilgenbusch *et al.* 2004).

The gene tree of individual loci for the four *Euphaea* species (*decorata*, *formosa*, *ornata* and *yayeyamana*) was reconstructed using haplotype data sets in MRBAYES, with prior settings and the parameters of nucleotide substitution models estimated using JMODELTEST. The chosen models for each locus were as follows: JC for *act*; HKY for *anon* and *mlc*; TrN+G for *arr*, *awd2* and *lop1*; TPM2uf+G for *cox2*; TrNef+G for *EF1 α* ; TPM3uf+I+G for *fer*; TPM3+G for *ITS*; HKY+I+G for *nad5*; F81+I for *sdhB*. Two independent Bayesian analyses, each with four Markov chains, were run for 1×10^7 generations, with parameters sampled every 1×10^3 generations, and the first 25% (2500 trees) being discarded as burn-in. The convergence of runs was determined by examining the average split frequencies. A 50% majority consensus tree was reconstructed from the remaining trees in PAUP* to calculate the BPP.

Testing assumptions of the IM models

The IM model of species divergence (Nielsen & Wakeley 2001; Hey & Nielsen 2004) was used to estimate the

demographic parameters of population sizes, gene flow and divergence time. The IM models (IM, IMA and IMA2) were derived under several simplifying assumptions about the species' history of divergence and the genetic data, including selective neutrality of the loci, no intralocus recombination, mutation following infinite sites (IS) or the Hasegawa–Kishino–Yano (HKY) model, no population structure within species, demographic stability of diverging populations, and no genetic contribution from un-sampled population or species (Hey & Nielsen 2004, 2007; Pinho & Hey 2010). Strasburg & Rieseberg (2010) demonstrated that parameter estimates of IMA are generally robust for small to moderate violations of model assumptions. In particular, the results of their simulations indicated that a substantial population structure within diverging species has little effect on parameter estimates of IMA, and most parameter estimates are robust for considerable levels of recombination with loci containing nonrecombining blocks (Strasburg & Rieseberg 2010). In contrast, deviations from mutation models (Strasburg & Rieseberg 2010), ancestral population structure and variation in gene flow through time (Becquet & Przeworski 2009) can have a significant effect on most parameter estimates of IMA. We tested four assumptions of the IM models (i.e. neutrality, intralocus recombination, mutation model and population structure) to assess their potential effect on the results when model violations were detected.

The assumption of neutrality was tested using Tajima's *D* (Tajima 1989) and Fay and Wu's *H* (Fay & Wu 2000) in DNASP. The significance level of the test statistics was evaluated by performing 1000 coalescent simulations, with simultaneous estimation of recombination rate, *R* ($4Nr$, Hudson 1987). Additionally, the multilocus Hudson–Kreitman–Aguadé (HKA) test (Hudson *et al.* 1987) was conducted to examine the level of sequence polymorphism within, and sequence divergence between, species in the HKA program (<http://genfaculty.rutgers.edu/hey/software#HKA>), with significance determined on the basis of 10 000 coalescent simulations. The minimum number of recombination events (*Rm*) of each nuclear locus was estimated using the four-gamete test (Hudson & Kaplan 1985) in DNASP. When a high level of recombination within loci was detected (>5 recombination events), the IMgc (Woerner *et al.* 2007) program was used to extract maximal blocks of nonrecombining sequences for the individual locus. These nonrecombining sequences were subjected to additional IMA2 analyses. To select proper models of nucleotide substitution for IMA2 analyses, we evaluated the assumption of the infinite site (IS) model that each mutation occurs at a unique site (Kimura 1969) for each locus in IMA2. Of all sampled loci, *ITS* and *awd2* (*E. formosa* & *E. yayeyamana*) and *act* and *sdhB* (*E. decorata* &

E. ornata) fit the IS model. For the remaining loci, the HKY model was applied. The genetic structure within and among species was examined using a Bayesian clustering analysis of the admixture model implemented in STRUCTURE (v. 2.3.3, Pritchard *et al.* 2000; Hubisz *et al.* 2009). The data set analysed contains haplotype sequences of all sampled loci, with all individuals being represented by a single allele. The haplotypes of heterozygous sequences were determined in DNASP and randomly chosen for the analysis. The estimated number of clusters (K) was set to range from one to six for each species and for all individuals, to explore their optimal values. Ten independent MCMC runs of 1×10^6 steps were performed for each number of K , with the first 1×10^5 steps being discarded as burn-in.

IMa2 analyses

The IMa2 program (Hey & Nielsen 2004, 2007; Hey 2010) was used to examine the fit between the genetic data and the IM models in the two pairs of sibling *Euphaea* species. Six population genetic parameters were simultaneously estimated using an MCMC sampling approach. These parameters included the effective population size of the ancestral (θ_A) and two descendant populations (θ_1 and θ_2) ($\theta = 4N_e\mu$, where N_e is the effective population size and μ is the geometric mean of the mutation rates of all loci), two-directional migration rates (m_1 and m_2 , migration per locus per generation) and the divergence time since the split of the two descendant populations (t , in generations). Three to five short preliminary MCMC runs were carried out to define the boundaries of priors of the six parameters and to optimize the heating scheme of the Markov chains. Twenty Metropolis-coupled chains were run simultaneously, as recommended in the IMa2 manual for medium size data sets (<15 loci). A geometric heating scheme under the Metropolis-coupling criterion was used to improve mixing and convergence of the MCMC runs. The β value of the heating scheme, the degree of heat and the swapping rate between chains was kept to between 0.890 and 0.997 to achieve a moderate heating and a high swapping rate for the coupled chains, with the heating terms h_1 and h_2 set to 0.96 and 0.90, respectively. Three independent long runs with different random number seeds were conducted for each data set, with MCMC sampling restarted when the swapping rate between coupled chains was lower than 5%. Upper boundaries of the prior distributions for each parameter were set on the basis of the results of preliminary runs and were >10 times higher than those of the preliminary runs. These long runs were continued until the trend-line plots gave no visible trends and the

lowest ESS among the parameters was at least above 50, to ensure adequate mixing and convergence. Before reaching the plateau, all sampled genealogies of the long runs were discarded as burn-in. A total of 3×10^5 genealogies were saved after completion of the three long runs and used to calculate parameter values and likelihood ratio tests of the nested speciation models (Hey & Nielsen 2007).

To convert parameter estimates of population size and divergence time in IMa2 into demographic units, a revised mutation rate of mitochondrial cytochrome oxidase I (*cox1*) in insects (1.77×10^{-8} mutation/site/year, Papadopoulou *et al.* 2010) and a synonymous substitution rate of nuclear *Arrestin* (*arrB*) estimated in *Drosophila melanogaster-obscura* groups (1.2×10^{-8} mutation/site/year, Moriyama & Gojobori 1992) were applied to *cox2* and *arr* gene fragments, respectively. The mutation rates of the other loci were determined by calculating the ratio of the net genetic distance between sibling *Euphaea* species under the Tamura–Nei model at mitochondrial *nad5* vs. that at *cox2*, and at each nuclear locus vs. that at *arr*. The mutation rates of each locus were calculated separately for the two species pairs. The geometric means of mutation rates for the two sister pairs (3.58×10^{-6} /locus/year for *E. formosa* vs. *E. yayeyamana*; 3.32×10^{-6} /locus/year for *E. decorata* vs. *E. ornata*) were calculated from all 12 loci and used to convert the population size ($\theta = 4N\mu$) and divergence time ($t = T\mu$) into number of individuals (N) and years (T) (Hey & Nielsen 2004). A generation time of 1 year was assumed for these damselflies on the basis of available field observations (Dudgeon 1989; Hayashi 1990; Huang & Lin 2011).

Results

Species phylogeny and gene trees

A sequence alignment of 2124 bp (*cox2*: 705 bp; *arr*: 1419 bp) was obtained for ten *Euphaea* species and the outgroup, *Bayadera brevicauda*. This combined data set contained 323 (*cox2*: 170; *arr*: 153) parsimony informative sites. The MP, ML and Bayesian analyses of the combined data resulted in well-resolved species relationships and comparable tree topologies (Fig. S1, Supporting information). All trees recovered the two sister groups, *Euphaea formosa*/*E. yayeyamana* and *E. decorata*/*E. ornata*, with strong branch supports (Fig. S1, Supporting information). The species tree estimated with *BEAST recovered the same sister species pairs and supported each of the four *Euphaea* species as a reciprocally monophyletic group, with high posterior probability (Fig. 2). The genealogies of each locus showed no haplotypes shared between the two sister species pairs (Fig. 3). No haplotype sharing between sibling species

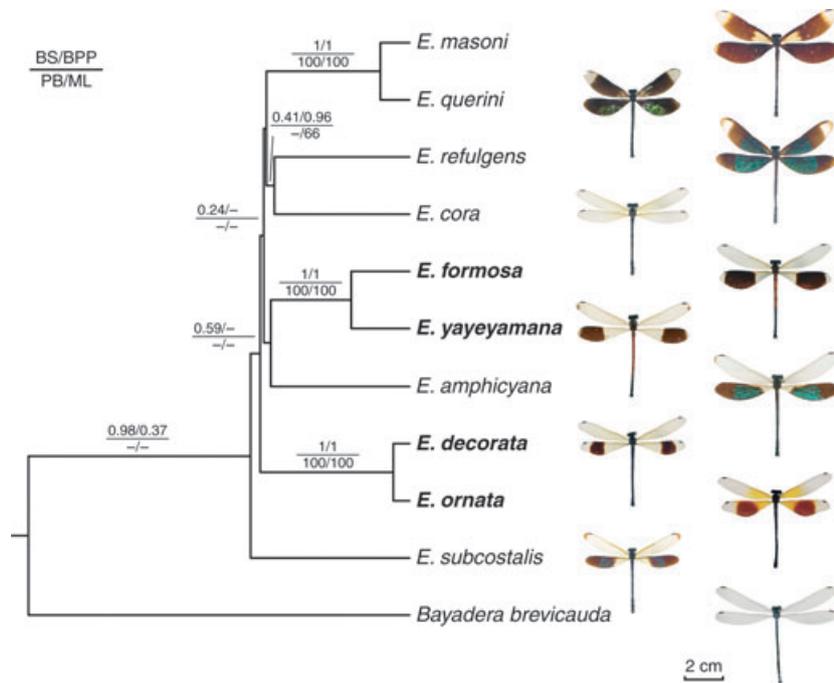


Fig. 2 Species phylogeny of *Euphaea* damselflies jointly estimated with *cox2* and *arr* genes in *BEAST. Numbers near the nodes are branch support values (BS, Bayesian posterior probabilities of species tree; BPP, Bayesian posterior probabilities; PB, parsimony bootstrapping; LB, likelihood bootstrapping).

within a species pair was detected at *ITS*, *cox2*, *arr* or *sdhB*. In contrast, *awd2*, *mlc* and *anon* showed extensive shared polymorphism between sibling species. In *nad5* genealogy, one haplotype from *E. formosa* was distributed at the tip of the nested haplotypes from *E. yayeyamana*, demonstrating the possibility of recent mitochondrial introgression from *E. formosa* to *E. yayeyamana*. Overall, the level of haplotype sharing of all loci between *E. formosa* and *E. yayeyamana* was higher than that between *E. decorata* and *E. ornata*.

Genetic diversity and model assumptions of IMA2

A total of 9659 (*E. formosa* & *E. yayeyamana*) and 9748 (*E. decorata* & *E. ornata*) bp from 12 loci (366–1381 bp per locus) were sequenced for each of 14–31 individuals from the four *Euphaea* species (Table 1). The levels of polymorphism were extensive in most loci, with average numbers of polymorphic sites (*S*) per locus ranging from 18 in *E. yayeyamana* to 33 in *E. formosa*. The haplotype diversity (H_d) was high for all loci (mean = 0.807–0.866) except *act* (0.489–0.591). The overall H_d was similar among the four *Euphaea* species, whereas the overall nucleotide diversity (π) was higher in *E. formosa* than the other three species. The net genetic distance per site showed considerable variation among loci, and the average divergence between *E. formosa* and *E. yayeyamana* (0.014) was approximately two (5–9 in mitochon-

drial loci) times greater than that between *E. decorata* and *E. ornata* (0.007) (Table S2, Supporting information). The *Fst*-based estimates of migration rates (*Nm*) indicated wide variation in levels of gene flow among loci, ranging from 0.03 (*ITS*) to 3.86 (*awd2*) between *E. formosa* and *E. yayeyamana* and from 0 (*act*) to 2.37 (*awd2*) between *E. decorata* and *E. ornata* (Table S2, Supporting information).

Considerable numbers of minimum recombination events ($R_m \geq 5$) were detected by four-gamete tests in *mlc* (503 bp), *lop1* (596 bp) and *anon* (147 bp) of *E. formosa*, *arr* (929 bp) of *E. yayeyamana*, *arr* (599 bp), *fer* (928 bp) and *anon* (254 bp) of *E. decorata*, and *arr* (599 bp) and *lop1* (755 bp) of *E. ornata*, whereas intralocus recombination in *ITS*, *act*, *EF1 α* and *sdhB* was low (Table 1). The maximum nonrecombined regions (7780 bp in *E. formosa* and *E. yayeyamana*; 7921 bp in *E. decorata* and *E. ornata*) were extracted from all loci and subjected to additional IMA2 analyses to exclude the influence of recombination. The results of multilocus HKA tests demonstrated no significant departure from neutral expectation for all loci (*E. formosa* & *E. yayeyamana*, sum of deviations = 25.118, d.f. = 22, $P = 0.291$; *E. decorata* & *E. ornata*, sum of deviations = 11.294, d.f. = 22, $P = 0.971$). The multilocus Tajima's *D* did not significantly deviate from neutral expectation (*E. formosa*, $D = -0.462$, $P = 0.097$; *E. yayeyamana*, $D = 0.097$, $P = 0.660$; *E. decorata*, $D = 0.006$,

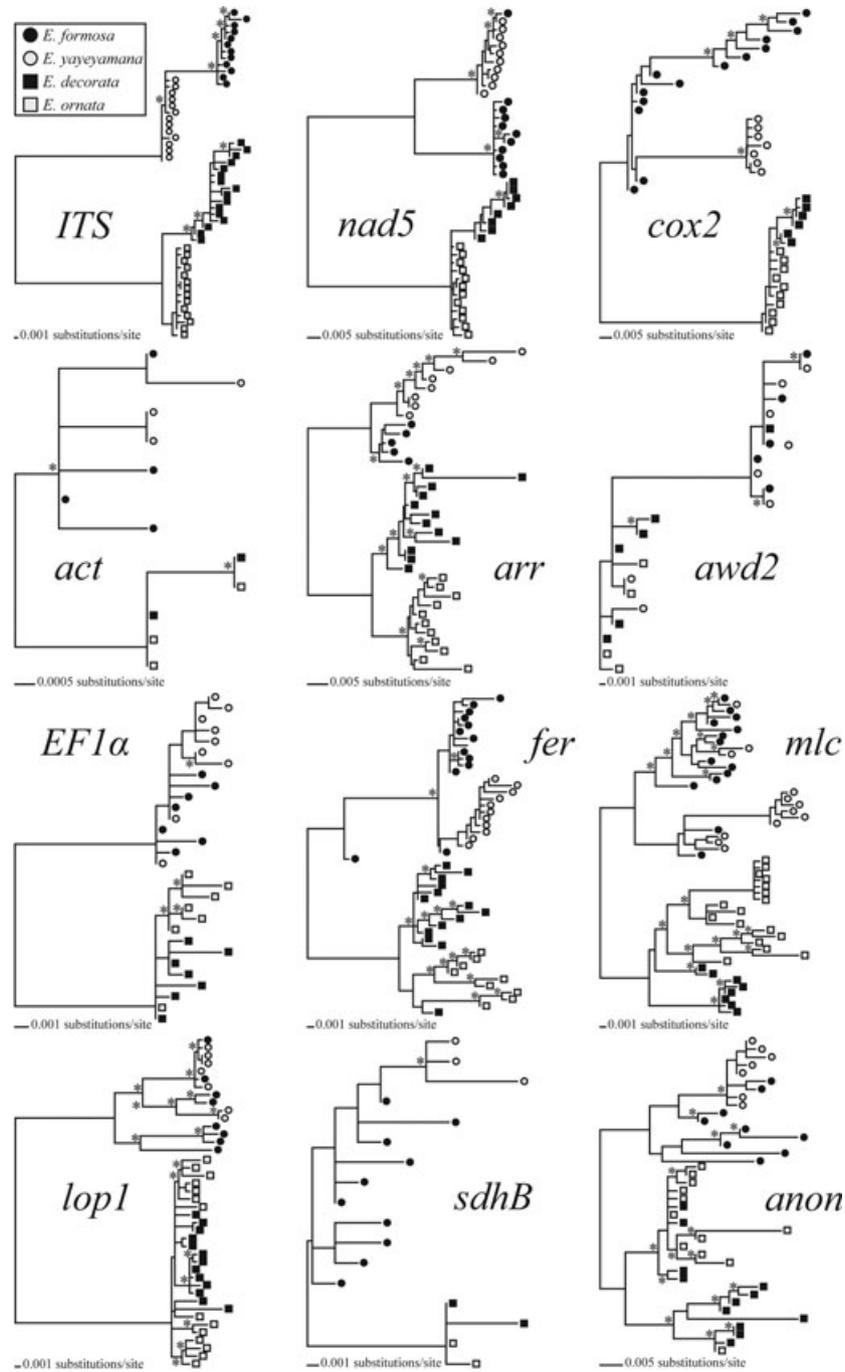


Fig. 3 Gene trees for the four *Euphaea* damselflies based on each of the 12 loci used in this study. The asterisks by the branches indicate nodes supported by Bayesian posterior probabilities above 90%.

$P = 0.640$; *E. ornata*, $D = -0.451$, $P = 0.077$). Tajima's D tests of individual loci demonstrated that most loci had negative values of D , but were not significantly different from zero, except for *nad5* in *E. formosa* ($D = -2.192$, $P < 0.001$) and *E. ornata* ($D = -1.839$, $P = 0.006$) (Table 1). Fay and Wu's H test, which is less sensitive to demographic fluctuation, indicated a significant

departure from neutrality in *nad5* of *E. formosa* and *E. yayeyamana*, *fer* of *E. yayeyamana*, *mlc* of *E. formosa*, *E. yayeyamana* and *E. decorata* and *anon* of *E. ornata* (Table 1). Therefore, additional IMA2 analyses were performed with the exclusion of loci potentially under selection (*nad5*, *fer* and *mlc* in *E. formosa* & *E. yayeyamana*, 6645 bp; *mlc* and *anon* in *E. decorata* & *E. ornata*,

Table 1 Polymorphism and summary statistics for the 12 loci analysed within each of the four *Euphaea* species

Locus	Species	Coding	Noncoding	<i>N</i>	<i>S</i>	<i>S_e</i>	<i>S_i</i>	<i>H_d</i>	π	<i>R_m</i>	<i>H</i>	<i>D</i>
<i>cox2</i>	<i>formosa</i>	500	0	19	42	42	0	0.947	0.0289	–	2.111	0.497
	<i>yayeyamana</i>	500	0	16	8	8	0	0.850	0.0032	–	–2.750	–1.225
	<i>decorata</i>	606	0	17	9	9	0	0.647	0.0053	–	0.191	0.756
	<i>ornata</i>	606	0	16	12	12	0	0.908	0.0035	–	1.800	–1.602
<i>nad5</i>	<i>formosa</i>	912	0	16	62	62	0	0.908	0.0101	–	–85.983**	–2.192**
	<i>yayeyamana</i>	912	0	16	16	16	0	0.925	0.0050	–	–6.833*	–0.202
	<i>decorata</i>	912	0	16	14	14	0	0.742	0.0058	–	–1.767	0.950
	<i>ornata</i>	912	0	15	17	17	0	0.943	0.0033	–	–0.343	–1.839*
<i>ITS</i>	<i>formosa</i>	666	0	19	15	15	0	0.942	0.0041	0	–	–
	<i>yayeyamana</i>	666	0	31	6	6	0	0.596	0.0015	0	–	–
	<i>decorata</i>	680	0	25	24	24	0	0.970	0.0073	1	–	–
	<i>ornata</i>	680	0	29	11	11	0	0.727	0.0016	0	–	–
<i>act</i>	<i>formosa</i>	445	0	19	3	3	3	0.591	0.0015	0	0.444	–0.607
	<i>yayeyamana</i>	445	0	18	1	1	1	0.503	0.0011	0	0.183	1.378
	<i>decorata</i>	445	0	18	1	1	1	0.503	0.0011	0	0.183	1.378
	<i>ornata</i>	445	0	24	1	1	1	0.489	0.0011	0	–0.326	1.391
<i>arr</i>	<i>formosa</i>	305	1029	15	25	5	20	0.838	0.0056	0	–2.590	–0.151
	<i>yayeyamana</i>	305	1029	19	51	5	46	0.883	0.0108	7	6.368	–0.154
	<i>decorata</i>	305	1016	21	64	16	48	0.952	0.0125	13	1.752	–0.363
	<i>ornata</i>	305	1016	20	49	10	39	0.947	0.0098	5	6.358	–0.329
<i>awd2</i>	<i>formosa</i>	154	212	19	5	1	4	0.789	0.0039	0	0.561	–0.084
	<i>yayeyamana</i>	154	212	17	16	2	14	0.809	0.0115	0	2.919	–0.695
	<i>decorata</i>	154	212	21	14	1	13	0.805	0.0095	1	1.214	–0.421
	<i>ornata</i>	154	212	19	12	2	10	0.813	0.0083	0	2.041	–0.705
<i>EF1α</i>	<i>formosa</i>	785	243	18	11	9	2	0.784	0.0018	0	0.601	–1.561
	<i>yayeyamana</i>	785	243	19	8	8	0	0.889	0.0020	2	1.076	–0.292
	<i>decorata</i>	782	243	21	10	8	2	0.814	0.0019	1	1.505	–1.027
	<i>ornata</i>	782	243	20	9	6	3	0.637	0.0014	0	–0.632	–1.441
<i>fer</i>	<i>formosa</i>	337	876	24	43	10	33	0.884	0.0059	3	–2.377	–1.576
	<i>yayeyamana</i>	337	876	25	18	5	13	0.900	0.0032	3	–9.930**	–0.705
	<i>decorata</i>	337	878	20	84	11	73	0.974	0.0120	7	–8.389	–1.624
	<i>ornata</i>	337	878	14	38	8	30	0.945	0.0139	2	2.022	1.699
<i>mlc</i>	<i>formosa</i>	193	696	22	53	4	49	0.952	0.0163	6	–9.333*	–0.313
	<i>yayeyamana</i>	193	696	20	40	1	39	0.953	0.0133	0	–12.294*	–0.412
	<i>decorata</i>	193	700	19	26	2	24	0.883	0.0085	1	–9.994*	–0.337
	<i>ornata</i>	193	700	24	53	3	50	0.975	0.0176	1	0.964	0.115
<i>lop1</i>	<i>formosa</i>	311	1070	18	85	4	81	0.922	0.0248	5	4.837	1.375
	<i>yayeyamana</i>	311	1070	16	38	0	38	0.875	0.0131	0	0.717	2.392*
	<i>decorata</i>	311	1057	23	47	5	42	0.960	0.0072	0	6.917	–0.920
	<i>ornata</i>	311	1057	21	52	5	47	0.967	0.0083	5	7.957	–0.890
<i>sdhB</i>	<i>formosa</i>	275	200	21	14	13	1	0.914	0.0080	1	–0.771	–0.344
	<i>yayeyamana</i>	275	200	17	6	5	1	0.750	0.0056	1	1.581	0.962
	<i>decorata</i>	258	200	14	3	2	1	0.495	0.0032	0	–1.187	1.753
	<i>ornata</i>	258	200	21	1	1	0	0.429	0.0009	0	0.257	0.959
<i>anon</i>	<i>formosa</i>	0	450	17	33	–	–	0.919	0.0228	6	0.816	–0.090
	<i>yayeyamana</i>	0	450	17	7	–	–	0.875	0.0048	1	–2.265	0.079
	<i>decorata</i>	0	459	20	35	–	–	0.953	0.0250	5	1.842	0.383
	<i>ornata</i>	0	459	24	32	–	–	0.902	0.0131	3	–10.601*	–1.121
Total/mean	<i>formosa</i>	4494	5165	227	32.6	–	–	0.866	0.011	21	–	0.021
	<i>yayeyamana</i>	4494	5165	231	17.9	–	–	0.817	0.006	14	–	0.525*
	<i>decorata</i>	4580	5168	235	27.6	–	–	0.808	0.008	29	–	0.595**
	<i>ornata</i>	4580	5168	247	23.9	–	–	0.807	0.007	16	–	–0.116

Summary statistics were calculated in DNASP v.5.0 (Librado & Rozas 2009).

N, number of sequences; *S*, number of polymorphic sites; *S_e*, number of polymorphic sites in exons; *S_i*, number of polymorphic sites in introns; *H_d*, haplotype diversity; π , nucleotide diversity; *R_m*, minimum number of recombination events; *H*, Fay and Wu's *H*; *D*, Tajima's *D*. **P* < 0.05; ***P* < 0.01.

8396 bp). STRUCTURE analyses revealed no significant genetic differentiation within each of the four *Euphaea* species. The most probable number of clusters for the individuals in each of the four species was one ($K = 1$) (Fig. S2A, Supporting information). The clustering analysis of all individuals indicated the presence of four genetic clusters ($K = 4$), corresponding to delineation of the four *Euphaea* species (Fig. S2B,C, Supporting information).

Estimation of demographic parameters and selection of speciation model using IMA2

Multiple runs of IMA2 simulation using the full data set resulted in posterior probability distributions of the

population size (θ), divergence time (t) and migration rate (m), with unambiguous peaks and bounds within the ranges of prior values (Fig. 4). The ESS values for the splitting time parameters in different IMA2 runs were higher than the recommended 50 for the proper mixing of runs (481–3783 in *E. formosa* & *E. yayeyamana*; 255–2555 in *E. decorata* & *E. ornata*). However, the upper boundary of the posterior distribution of the divergence time parameter for *E. formosa* and *E. yayeyamana* did not reach zero over a wide range of priors and resulted in a wide posterior density interval (Fig. 4B). The peak of the posterior distribution of ancestral effective population size for *E. formosa* and *E. yayeyamana* was less pronounced across a broad range of priors (Fig. 4A).

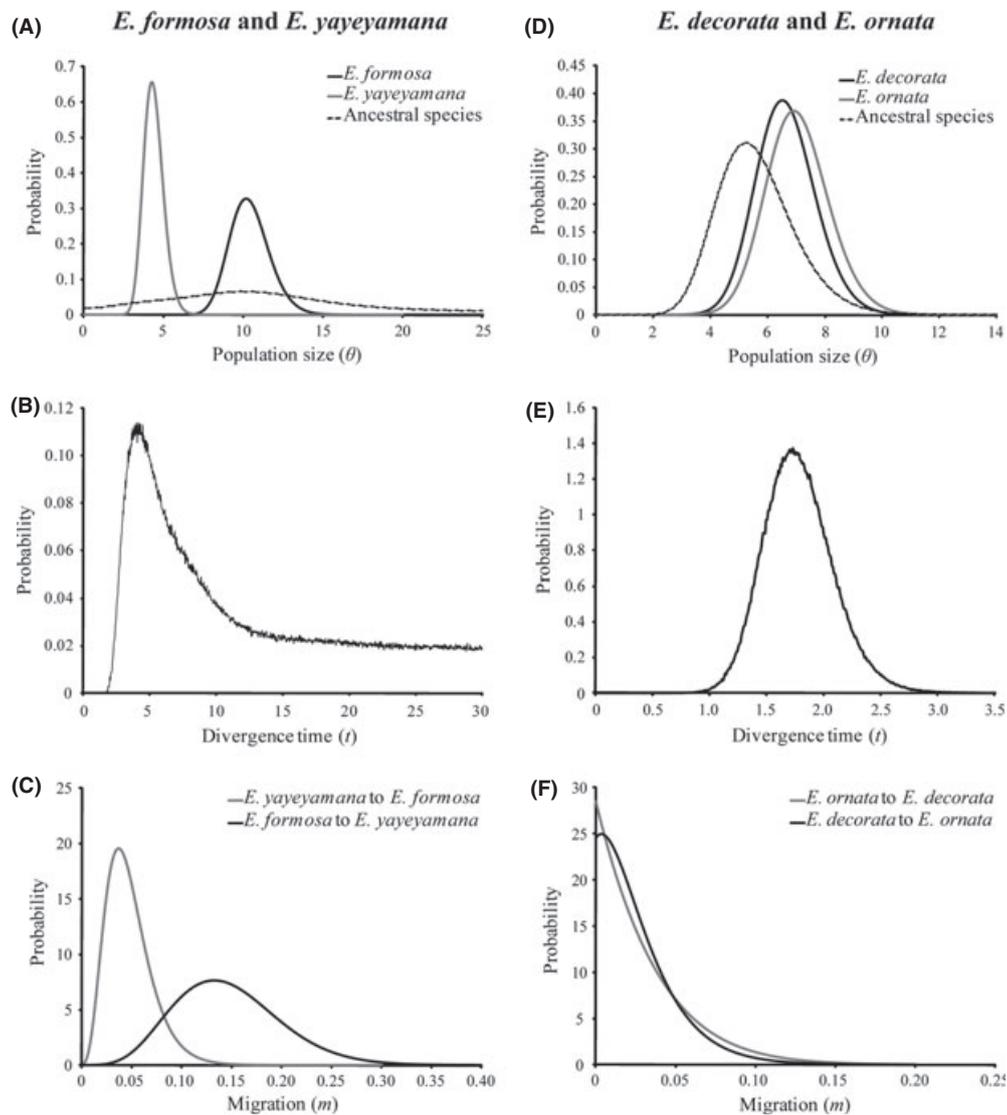


Fig. 4 The marginal posterior probability density distributions of model parameters (scaled by the mutation rate) estimated using the full data set in IMA2. Curves are shown for the analysis with *Euphaea formosa* and *E. yayeyamana* (A), (B) and (C), and for *E. decorata* and *E. ornata* (D), (E) and (F).

The maximum-likelihood estimate (MLE) and 95% highest posterior density (HPD) of parameters in mutational and demographic units are shown in Table 2. The IMA2 analysis of the full data set demonstrated an approximately equal effective population size for *E. formosa* ($N_f = 0.711$ million, HPD: 0.561–0.903) and the ancestral species ($N_A = 0.701$ million, HPD: 0–2.408), and a smaller effective population size for *E. yayeyamana* ($N_y = 0.298$ million, HPD: 0.224–0.393) (Table 2). The effective population sizes of *E. decorata* ($N_d = 0.489$ million, HPD: 0.352–0.661) and *E. ornata* ($N_o = 0.521$ million, HPD: 0.374–0.698) were approximately equal (Table 2). The ancestral species of *E. decorata* and *E. ornata* had a slightly smaller effective population size ($N_A = 0.397$ million, HPD: 0.237–0.625) than those of its descendants.

The divergence time for *E. formosa* and *E. yayeyamana* was 1.145 million years ago (Ma) (HPD: 0.642–8.334), in the Calabrian stage of the Lower Pleistocene period, which was approximately twofold earlier than the split between *E. decorata* and *E. ornata* at 0.511 Ma (HPD: 0.357–0.726), in the Ionian stage of the Middle Pleistocene (Table 2). Similar results were obtained by applying the standard insect mitochondrial clock (*E. formosa*/*E. yayeyamana*: 1.198 Ma, HPD: 0.671–8.718; *E. decorata*/*E. ornata*: 0.535 Ma, HPD: 0.374–0.760). Although analysis of the full data set produced a posterior distribution with a long upper boundary for the divergence time of *E. formosa* and *E. yayeyamana* (Fig. 4B), the sharp peaks of the posterior probability distribution and MLE for divergence time in both species pairs provided sufficient information to suggest

that *E. formosa* and *E. yayeyamana* diverged earlier than *E. decorata* and *E. ornata* (Fig. 4B,E, Table 2).

Analyses of the full data set indicated nonzero and significant bidirectional gene flow between *E. formosa* to *E. yayeyamana* ($m_{fy} = 0.133$, $\chi^2 = 38.192$, $P < 0.001$; $m_{yf} = 0.037$, $\chi^2 = 30.821$, $P < 0.001$) (Fig. 4C, Table 2). The migration rate per generation from *E. formosa* to *E. yayeyamana* ($2N_f M_{fy}$) was 0.675 (HPD: 0.200–1.623), which was approximately nine times higher than the reverse migration from *E. yayeyamana* to *E. formosa* ($2N_y M_{yf}$) at 0.079 (HPD: 0.015–0.263) (Table 2). The number of migration events for each locus indicated that all twelve loci showed evidence of migration from *E. formosa* to *E. yayeyamana* (1–5 events), whereas the migration event from *E. yayeyamana* to *E. formosa* was only found in four loci (*nad5*, *mlc*, *lop1* and *anon*, 1–2 events) (Table S3, Supporting information). The distribution of the mean times of migration events for each locus was broad (m_{fy} , $t = 0.02$ – 4.04 ; m_{yf} , $t = 0.32$ – 1.85) (Table S3, Supporting information). In contrast, gene flow between *E. decorata* and *E. ornata* was inferred to be negligible (MLE of m_{do} & $m_{od} < 0.001$) (Table 2), with the lower 95% HPD of both migration parameters including zero (Fig. 4F), and only one migration event detected from *E. decorata* to *E. ornata* (*anon*, Table S3, Supporting information).

Table 3 shows the likelihood ratio statistics for a series of nested speciation models applied to the full data set. Between *E. formosa* and *E. yayeyamana*, only the model of an equal effective population size for *E. formosa* (θ_f) and its ancestral species (θ_A) was not rejected ($P = 0.189$, Table 3). All remaining models, in which

Table 2 Maximum-likelihood estimates (MLE) and the 95% highest posterior density (HPD) intervals of demographic parameters estimated using the full data set in IMA2

Full data set	θ_f	θ_y	θ_A	m_{fy}	m_{yf}	t	N_f	N_y	N_A	$2N_f M_{fy}$	$2N_y M_{yf}$	T
<i>Euphaea formosa</i> and <i>E. yayeyamana</i>												
MLE	10.180	4.270	10.030	0.1326	0.0367	4.095	0.711	0.298	0.701	0.6749	0.0785	1.145
HPD95Lo	8.025	3.210	0	0.0498	0.0092	2.295	0.561	0.224	0	0.1998	0.0147	0.642
HPD95Hi	12.930	5.630	29.310	0.2510	0.0935	29.81	0.903	0.393	2.048	1.6227	0.2631	8.334
Full data set	θ_d	θ_o	θ_A	m_{do}	m_{od}	t	N_d	N_o	N_A	$2N_d M_{do}$	$2N_o M_{od}$	T
<i>E. decorata</i> and <i>E. ornata</i>												
MLE	6.492	6.912	5.265	0.0009	0.0001	1.696	0.489	0.521	0.397	0.0029	0.0003	0.511
HPD95Lo	4.673	4.963	3.141	0	0	1.184	0.352	0.374	0.237	0	0	0.357
HPD95Hi	8.768	9.262	8.289	0.0735	0.0877	2.408	0.661	0.698	0.625	0.3222	0.4061	0.726

θ_f , N_f , effective population size of *E. formosa*; θ_y , N_y , effective population size of *E. yayeyamana*; θ_d , N_d , effective population size of *E. decorata*; θ_o , N_o , effective population size of *E. ornata*; θ_A , N_A , effective population size of ancestral species; m_{fy} , $2N_f M_{fy}$, population migration rate from *E. formosa* to *E. yayeyamana*; m_{yf} , $2N_y M_{yf}$, population migration rate from *E. yayeyamana* to *E. formosa*; m_{do} , $2N_d M_{do}$, population migration rate from *E. decorata* to *E. ornata*; m_{od} , $2N_o M_{od}$, population migration rate from *E. ornata* to *E. decorata*; t , T , divergence time.

θ_f , θ_y , θ_d , θ_o , θ_A , m_{fy} , m_{yf} , m_{do} , m_{od} and t are scaled by the mutation rate; N_f , N_y , N_d , N_o , $2N_f M_{fy}$, $2N_y M_{yf}$, $2N_d M_{do}$, $2N_o M_{od}$, and T are in demographic units ($\times 10^6$ individuals or years).

Table 3 Tests of nested demographic models for the two *Euphaea* species pairs based on the full data set

<i>E. formosa</i> and <i>E. yayeyamana</i>				<i>E. decorata</i> and <i>E. ornata</i>			
Model	d.f.	-2LLR*	<i>P</i>	Model	d.f.	-2LLR*	<i>P</i>
$m_{fy} = m_{yf}$	1	35.69	***	$m_{do} = m_{od}$	1	0.001	1.000
$m_{yf} = 0$	1	110.10	***	$m_{od} = 0$	1	0.001	0.500
$m_{fy} = 0$	1	272.20	***	$m_{do} = 0$	1	0.001	1.000
$\theta_f = \theta_y$	1	132.50	***	$\theta_d = \theta_o$	1	0.587	0.444
$\theta_f = \theta_y, m_{fy} = m_{yf}$	2	137.90	***	$\theta_d = \theta_o, m_{do} = m_{od}$	2	0.587	0.745
$\theta_f = \theta_y, m_{yf} = 0$	2	245.20	***	$\theta_d = \theta_o, m_{od} = 0$	2	0.587	0.373
$\theta_f = \theta_y, m_{fy} = 0$	2	316.10	***	$\theta_d = \theta_o, m_{do} = 0$	2	0.587	0.373
$\theta_f = \theta_A$	1	1.73	0.189	$\theta_d = \theta_A$	1	0.006	0.940
$\theta_f = \theta_A, m_{fy} = m_{yf}$	2	40.24	***	$\theta_d = \theta_A, m_{do} = m_{od}$	2	0.006	0.997
$\theta_f = \theta_A, m_{yf} = 0$	2	111.80	***	$\theta_d = \theta_A, m_{od} = 0$	2	0.006	0.499
$\theta_f = \theta_A, m_{fy} = 0$	2	273.00	***	$\theta_d = \theta_A, m_{do} = 0$	2	0.006	0.499
$\theta_y = \theta_A$	1	27.99	***	$\theta_o = \theta_A$	1	0.219	0.640
$\theta_y = \theta_A, m_{fy} = m_{yf}$	2	73.98	***	$\theta_o = \theta_A, m_{do} = m_{od}$	2	0.219	0.896
$\theta_y = \theta_A, m_{yf} = 0$	2	138.10	***	$\theta_o = \theta_A, m_{od} = 0$	2	0.219	0.448
$\theta_y = \theta_A, m_{fy} = 0$	2	310.00	***	$\theta_o = \theta_A, m_{do} = 0$	2	0.219	0.448
$\theta_f = \theta_y = \theta_A$	2	141.40	***	$\theta_d = \theta_o = \theta_A$	2	0.634	0.728
$\theta_f = \theta_y = \theta_A, m_{fy} = m_{yf}$	3	156.10	***	$\theta_d = \theta_o = \theta_A, m_{do} = m_{od}$	3	0.634	0.889
$\theta_f = \theta_y = \theta_A, m_{yf} = 0$	3	251.50	***	$\theta_d = \theta_o = \theta_A, m_{od} = 0$	3	0.634	0.445
$\theta_f = \theta_y = \theta_A, m_{fy} = 0$	3	329.30	***	$\theta_d = \theta_o = \theta_A, m_{do} = 0$	3	0.634	0.445

*The log likelihood ratio (LLR) statistics and associated *P* values of the nested models calculated in IMA2; -2LLR follows a χ^2 distribution with the degree of freedom (d.f.) being the difference in the number of parameters between nested models.

P* < 0.01; *P* < 0.001.

the two migration rates are zero or equal to each other ($m_{fy} = 0$; $m_{yf} = 0$; $m_{fy} = m_{yf}$) and the ones in which the two effective population sizes ($\theta_f = \theta_y$) are equal were rejected. In contrast, the likelihood ratio statistic between *E. decorata* and *E. ornata* was not significant, so the nested models of zero migration ($m_{do} = 0$; $m_{od} = 0$) or equal population sizes ($\theta_d = \theta_o$) could not be rejected (Table 3).

The posterior probability distributions and likelihood estimates of parameters and nested models obtained from the three smaller data sets (no recombined blocks, no non-neutral loci, and no recombined blocks and non-neutral loci) revealed similar patterns to the full data set (summarized in Tables S4–S5 and Figs S3–S5, Supporting information). In all three data sets, *E. formosa* and *E. yayeyamana* were estimated to have unequal population sizes and nonzero asymmetric gene flows (Figs S3–S5A,C, Supporting information), while *E. decorata* and *E. ornata* were inferred to have equal population sizes and close to zero gene flows (Figs S3–S5D,F, Supporting information). However, these analyses generally resulted in estimates of smaller population sizes, more recent divergence times, and higher migration rates for both species pairs when compared to the full data set. The estimates of the divergence time parameter (*t*) for *E. formosa* and *E. yayeyamana* failed to converge in all three data sets (Figs. S3–S5B, Support-

ing information), and resulted in a large HPD interval and a high upper boundary, as reported in other studies using IM models (e.g. Kotlik *et al.* 2008; Nadachowska & Babik 2009; Morgan *et al.* 2010; Wachowiak *et al.* 2011). This result suggests that the information in these smaller data sets may not be sufficient to estimate the divergence time, a parameter that often requires a large number of unlinked loci to obtain an unambiguous probability distribution and a narrow credibility interval (Won & Hey 2005; Hey 2010).

Discussion

Species status, genetic differentiation and phylogeny

Species status, genetic variation and the phylogenetic relationship between *Euphaea formosa* and *E. yayeyamana* have been studied using sequences from mitochondrial *cox2* and the nuclear *ITS* gene (Huang & Lin 2011; Lee & Lin 2012). The major constraint of these earlier studies was the limited number of loci and taxa employed, as the stochastic nature of single-locus coalescence can mislead the interpretation. Based on the coalescent species tree reconstruction of *cox2* and *arr* genes, the present study supported the previous findings of distinct 'phylogenetic species' (Cracraft 1983) for *E. formosa* and *E. yayeyamana* and demonstrated

that the four *Euphea* species are each monophyletic lineages, a prerequisite for applying IM models to test the mode of species formation rather than population divergence (Hey 2006; Pinho & Hey 2010). Genetic clustering analyses of the multilocus data set also demonstrated substantial genetic differentiation among the species, and few structured populations within the species that further supported valid genotypic species (Mallet 1995) for all four *Euphaea* species. The results of distinct phylogenetic and genotypic species statuses for *E. formosa* and *E. yayeyamana* were consistent with earlier conclusions of separate 'morphological/phenetic species' (Sokal & Crovello 1970), as characterized by distinctive, nonoverlapping male wing shapes (Lee & Lin 2012).

In contrast to the well-characterized phylogenetic and genotypic species status for the four *Euphaea* damselflies, gene trees inferred from *awd2*, *mlc* and *anon* revealed extensive shared polymorphism and multiple persisting polyphyletic or paraphyletic haplotype lineages between the sibling *Euphaea* species, whereas only the *ITS*, *cox2*, *arr* and *sdhB* gene trees indicated reciprocal monophyly of the four *Euphaea* species. The inconsistency of species relationships among multiple gene trees suggested that these four damselflies have not yet become complete genealogical species (Baum & Shaw 1995) across the sampled loci. Discordance between gene trees and species trees can be caused by the stochastic coalescent process, interspecific hybridization/gene flow, or incomplete lineage sorting, which is to be expected for lineages during their early stages of speciation (Maddison 1997). No haplotype was shared between the two sibling *Euphaea* species pairs for all sampled loci, demonstrating that these two sibling species pairs have been diverged for a long time, allowing complete lineage sorting.

Timing of speciation in *Euphaea* damselflies

Analyses of the multilocus data set support the recent divergence of *E. formosa*, *E. yayeyamana*, *E. decorata* and *E. ornata* within the last 1 Myr. Therefore, two speciation events of sibling *Euphaea* species pairs date to the Middle to Lower Pleistocene period, during which the periodic connection and isolation of forest habitats through glacial land bridges between the Asian mainland and continental islands would have significantly affected species distribution and gene flow between diverging populations. The estimated divergence time for *E. formosa* and *E. yayeyamana*, based on the present data set (1.2 Ma, HPD: 0.64–8.33, Table 2), was approximately twofold more recent than the previous estimation based on the mitochondrial *cox2* gene (2.6 Ma, HPD: 0.483–6.216, Huang & Lin

2011). The results of coalescent simulations based on 12 loci indicated that the MLE of divergence time was conservative and associated with considerable uncertainty (Table 2, Fig. 4). Nevertheless, the inconsistency between the point estimates of divergence time using a single mitochondrial gene and multilocus data set might result partly from applying a faster mitochondrial mutation rate derived from time-calibrated biogeographical events (1.77×10^{-8} mutation/site/year, Papadopoulou *et al.* 2010), rather than the standard insect mitochondrial clock (1.15×10^{-8} mutation/site/year, Brower 1994) employed in the earlier study (Huang & Lin 2011).

Speciation with gene flow in *E. formosa* and *E. yayeyamana*

The IM analyses demonstrated significant bidirectional postdivergence gene flow between *E. formosa* and *E. yayeyamana* (Table 2, Fig. 4C). The level of gene flow was interpreted to be large from *E. formosa* to *E. yayeyamana* and moderate from *E. yayeyamana* to *E. formosa*, reflecting the passage of genes between species after selection against gene flow has acted (Pinho & Hey 2010). These results are therefore strongly inconsistent with a strictly allopatric speciation model (Table 3), which assumes complete termination of gene flow at the time of divergence between the two species. Alternatively, unambiguous signals of nonzero gene flow after species divergence favours the IM model to explain the origin of endemic *E. formosa* and *E. yayeyamana* on subtropical East Asian islands. In the strictly allopatric speciation model, after initial geographical isolation, mutation and genetic drift play a more important role than selection in driving population divergence, with reproductive isolation emerging as a byproduct of genetic divergence in isolated populations (Mayr 1954, 1963; Coyne & Orr 2004). In contrast, under the IM model of speciation, divergent selection is a major evolutionary force responsible for generating and maintaining population divergence in the face of homogenizing gene flow (Hey 2006; Nosil 2008). Theoretical models have suggested that strong ecological or sexual selection could have profound effects in speciation by increasing the rate of divergence of locally adapted populations (Gavrilets 2003, 2004). Therefore, factors, such as local ecological selection of prey availability and resource allocation (Hayashi 1990; Lee & Lin 2012), and sexual selection resulting from territorial competition for mates (Lee & Lin 2012), may have promoted divergence of these two insular *Euphaea*, even when land bridges were connecting their habitats between Taiwan and the Yaeyama islands during Pleistocene glaciated periods.

Asymmetric gene flow and peripatric speciation

The estimated gene flow is much greater from *E. formosa* to *E. yayeyamana* ($2N_f M_{fy} = 0.675$) than in the opposite direction ($2N_y M_{yf} = 0.079$). This asymmetric gene flow is commonly observed in ecologically diverged species (e.g. Morgan *et al.* 2010), consistent with a major role of divergent ecological adaptation in the speciation of these two damselflies. The asymmetry of gene flow can have several possible explanations. First, hybrid zone theory predicts that the migrant flow in natural populations tends to be higher from densely populated to sparsely populated areas owing to asymmetry of genetic barriers (Lenormand 2002). The asymmetric gene flow may be partly explained by a much larger ($N_f = 0.711$ million) and denser *E. formosa* population than that of *E. yayeyamana* ($N_y = 0.298$ million) (Hayashi 1990; Huang & Lin 2011). Second, greater interspecific gene flow is also expected from the species with higher fitness into the species with lower fitness as a consequence of asymmetric hybridization. Although the relative fitness of *E. formosa* and *E. yayeyamana* in each other's habitats is unknown, the larger *E. formosa*, with its broader hind wings for superior aerial manoeuvring ability, may have attained higher fitness through selective advantage in the territorial competition for mates (Lee & Lin 2012). This could have resulted in a trend towards greater gene flow from *E. formosa* to *E. yayeyamana* when the two species were in sympatry. Finally, hybrid zone theory suggests that moving hybrid zones are expected to leave a signature of asymmetric introgression in the direction of displacement of the hybrid zone (Buggs 2007). Our inference of asymmetric gene flow would therefore be in accordance with an eastward displacement of the hybrid zone from 'mainland' *E. formosa* into 'peripheral' *E. yayeyamana*. Nevertheless, distinguishing between these alternative explanations requires further experimentation.

Speciation without gene flow in E. decorata and E. ornata

Compared with *E. formosa* and *E. yayeyamana*, the post-divergence gene flow between *E. decorata* and *E. ornata* was not significantly different from zero to reject the strictly allopatric speciation model. The level of DNA polymorphism and effective population size of *E. decorata* ($\theta_d = 6.492$, $N_d = 0.489$ million), *E. ornata* ($\theta_o = 6.912$, $N_o = 0.521$ million) and their ancestral species ($\theta_A = 6.265$, $N_A = 0.397$ million) were inferred to be approximately equal, consistent with the prediction of a standard vicariant model of allopatric speciation (Mayr 1954, 1963; Coyne & Orr 2004), in that the ancestral population was split into two geographically isolated

populations with no gene flow, and at the time of speciation each of the isolated populations resembled the ancestral one in the level of DNA polymorphism, and thus was an effective population size (Osada & Wu 2005). Therefore, these results support the hypothesis of prolonged allopatric isolation of these two *Euphaea* species throughout much of the late Pleistocene, without periodic secondary contacts and consequent genetic exchange. Therefore, geographical isolation is probably to have played a more crucial role in the speciation of *E. decorata* and *E. ornata*. This allopatric scenario could have occurred as the rising sea level of the warm interglacials divided the ancestral population into the Asian mainland and Hainan island without gene exchange, and was reinforced when tropical damselfly populations became restricted to isolated forest refugia during the cold and arid glacial periods.

Conclusion

Our findings of contrasting geographical speciation modes (strict allopatry vs. parapatry/peripatry with gene flow) between the two sibling *Euphaea* species pairs demonstrate that each damselfly lineage has undergone divergent evolutionary trajectories leading to species formation, even with an apparently similar life history, ecological characteristics and phylogenetic affinity. We suggest that the effect of Pleistocene climatic changes on the origin of *Euphaea* species of the East Asian islands is lineage-specific and may depend predominantly on the divergence time and latitude of the evolutionarily independent clades. These twofold earlier divergences could have allowed the older *Euphaea formosa* and *E. yayeyamana* to experience larger numbers of glacial cycles and provided additional opportunities for secondary contacts and consequent gene flow than that of younger *E. decorata* and *E. ornata*. Latitudinally, the tropical island of Hainan lies approximately 200 km south of Taiwan, and the Yaeyama islands are located near the Tropic of Cancer, with a predominately subtropical climate. Tropical East Asia at the lower latitudes, including Hainan, was characterized by a relatively mild Pleistocene climate (Pinot *et al.* 1999) than that of Taiwan and the Yaeyama islands at the higher latitudes. Therefore, the land bridge formation between the Asian mainland and Hainan during glacial periods may have been less extensive and subsequently promoted prolonged allopatric isolation of *E. decorata* and *E. ornata* through oceanic or habitat barriers. We conclude that the role of geographical barriers may be less important for the speciation of subtropical *E. formosa* and *E. yayeyamana*, where differential selection of natural or sexual environments is prominent in driving species divergence despite homogenizing gene

flow; whereas for tropical *E. decorata* and *E. ornata* at lower latitudes, allopatric isolation may well be a pivotal promoter of species formation.

Acknowledgements

We would like to thank Hilario Cahilog, Arvin Diesmos, Euichi Hirose, Jen-Pan Huang, Masako Izawa, Syun Kobayashi, Katsuyuki Kohno, Hidetoshi Ota, Pham Hong Thai, Ta Huy Thinh, Reagan Villanueva, Arnulito Viojan and Lin Zhu for their help collecting damselfly specimens. The Department of Environment and Natural Resources (DENR) of the Philippines and the Institute of Ecology and Biological Resources (IEBR) of Vietnam issued research and collecting permits to facilitate fieldwork. We also thank Wei-Hsin Chiu, Hurng-Yi Wang and Chao-Tung Yang for access to the high-performance computation facility, and Jo-Fan Wang for performing phylogenetic analyses. We are grateful to Jake Esselstyn, Rosemary Gillespie, Shou-Hsien Li, Si-Min Lin and Hurng-Yi Wang for discussion and comments on earlier versions of this manuscript. The National Science Council of Taiwan provided financial support for this work (NSC 97-2621-B-029-001, 99-2621-M-029-001).

References

- Arias CF, Muñoz AG, Jiggins CD, Mavárez J, Bermingham E, Linares M (2008) A hybrid zone provides evidence for incipient ecological speciation in *Heliconius* butterflies. *Molecular Ecology*, **17**, 4699–4712.
- Baum D, Shaw KL (1995) Genealogical perspectives on the species problem. In: *Experimental and Molecular Approaches to Plant Biosystematics* (eds Hoch PC and Stephenson AG), pp. 289–303. Missouri Botanical Garden, St. Louis, Missouri.
- Becquet C, Przeworski M (2009) Learning about modes of speciation by computational approaches. *Evolution*, **63**, 2547–2562.
- Berner D, Grandchamp A, Hendry A (2009) Variable progress toward ecological speciation in parapatry: stickleback across eight lake-stream transitions. *Evolution*, **63**, 1740–1753.
- Brower AVZ (1994) Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 6491–6495.
- Brumfield RT, Liu L, Lum DE, Edwards SV (2008) Comparison of species tree methods for reconstructing the phylogeny of bearded manakins (Aves: Pipridae, *Manacus*) from multilocus sequence data. *Systematic Biology*, **57**, 719–731.
- Buggs RJA (2007) Empirical study of hybrid zone movement. *Heredity*, **99**, 301–312.
- Bull V, Beltrán M, Jiggins CD, McMillan WO, Bermingham E, Mallet J (2006) Polyphyly and gene flow between non-sibling *Heliconius* species. *BMC Biology*, **4**, 11.
- Bybee SM, Ogden TH, Branham MA, Whiting MF (2008) Molecules, morphology and fossils: a comprehensive approach to odonate phylogeny and the evolution of the odonate wing. *Cladistics*, **24**, 477–514.
- Carstens BC, Knowles LL (2007) Shifting distributions and speciation: species divergence during rapid climate change. *Molecular Ecology*, **16**, 619–627.
- Corbet PS (1999) *Dragonflies: Behavior and Ecology of Odonata*. Cornell University Press, Ithaca, New York.
- Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates Inc, Sunderland, Massachusetts.
- Cracraft J (1983) Species concepts and speciation analysis. *Current Ornithology*, **1**, 159–187.
- Drummond AJ, Rambaut A (2007) BEAST: bayesian evolutionary analysis by 629 sampling trees. *BMC Evolutionary Biology*, **7**, 214.
- Dudgeon D (1989) Life cycle, production, microdistribution and diet of the damselfly *Euphaea decorata* (Odonata: Euphaeidae) in a Hong Kong forest stream. *Journal of Zoology*, **217**, 57–72.
- Eklblom R, Galindo J (2010) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, **107**, 1–15.
- Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. *Genetics*, **155**, 1405–1413.
- Gavrilets S (2003) Perspective: models of speciation: what have we learned in 40 years? *Evolution*, **57**, 2197–2215.
- Gavrilets S (2004) *Fitness Landscapes and the Origin of Species*. Princeton University Press, Princeton, New Jersey.
- Hayashi F (1990) Convergence of insular dwarfism in damselflies (*Euphaea*) and dobsonflies (*Protophermes*). *Freshwater Biology*, **23**, 219–231.
- Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, **27**, 570–580.
- Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.
- Hey J (2006) Recent advances in assessing gene flow between diverging populations and species. *Current Opinion in Genetics & Development*, **16**, 592–596.
- Hey J (2010) The divergence of chimpanzee species and subspecies as revealed in multi-population isolation-with-migration analyses. *Molecular Biology and Evolution*, **27**, 921–933.
- Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **167**, 747–760.
- Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 2785–2790.
- Hey J, Wakeley J (1997) A coalescent estimator of the population recombination rate. *Genetics*, **145**, 833–846.
- Hoorn C, Wesselingh FP, ter Steege H *et al.* (2010) Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science*, **330**, 927–931.
- Huang JP, Lin CP (2011) Lineage-specific late Pleistocene expansion of an endemic subtropical gossamer-wing damselfly, *Euphaea formosa*, in Taiwan. *BMC Evolutionary Biology*, **11**, 94.
- Hubisz M, Falush D, Stephens M, Pritchard J (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322–1332.
- Hudson RR (1987) Estimating the recombination parameter of a finite population model without selection. *Genetical Research*, **50**, 245–250.

- Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*, **111**, 147–164.
- Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics*, **116**, 153–159.
- Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. *Genetics*, **132**, 583–589.
- Huelsenbeck JP, Ronquist F (2001) MrBayes v. 3.12: bayesian inference of phylogeny. *Bioinformatics*, **17**, 754–755.
- Jordan S, Simon C, Polhemus D (2003) Molecular systematics and adaptive radiation of Hawaii's endemic damselfly genus *Megalagrion* (Odonata: Coenagrionidae). *Systematic Biology*, **52**, 89–109.
- Kimura M (1969) The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics*, **61**, 893–903.
- Knowles LL (2000) Tests of Pleistocene speciation in montane grasshoppers (Genus *Melanopus*) from the sky islands of western North America. *Evolution*, **54**, 1337–1348.
- Kotlik P, Markova S, Choleva L *et al.* (2008) Divergence with gene flow between Ponto-Caspian refugia in an anadromous cyprinid *Rutilus frisii* revealed by multiple gene phylogeography. *Molecular Ecology*, **17**, 1076–1088.
- Leaché A, Mulcahy D (2007) Phylogeny, divergence times and species limits of spiny lizards (*Sceloporus magister* species group) in western North American deserts and Baja California. *Molecular Ecology*, **16**, 5216–5233.
- Lee YH, Lin CP (2012) Morphometric and genetic differentiation of two sibling gossamer-wing damselflies, *Euphaea formosa* and *E. yayeyamana* (Insecta, Odonata, Euphaeidae), and adaptive trait divergence in subtropical East Asian islands. *Journal of Insect Science*, **12**, 53.
- Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology and Evolution*, **17**, 183–189.
- Li JW, Yeung CKL, Tsai PW *et al.* (2010) Rejecting strictly allopatric speciation on a continental island: prolonged postdivergence gene flow between Taiwan (*Leucodioptron taewanus*, Passeriformes Timaliidae) and Chinese (*L. canorum canorum*) hwameis. *Molecular Ecology*, **19**, 494–507.
- Librado P, Rozas J (2009) DnaSP v. 5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.
- Losos JB, Glor RE (2003) Phylogenetic comparative methods and the geography of speciation. *Trends in Ecology and Evolution*, **18**, 220–227.
- Maddison W (1997) Gene trees in species trees. *Systematic Biology*, **46**, 523–536.
- Mallet J (1995) A species definition for the modern synthesis. *Trends in Ecology and Evolution*, **10**, 294–299.
- Matsuki K, Lien JC (1978) Descriptions of the larvae of three families of Zygoptera breeding in the streams of Taiwan. *Tombo*, **21**, 15–26.
- Matsuki K, Lien JC (1984) The euphaeid fauna of Taiwan (Odonata). *Rhopalocerists' Magazine*, **7**, 15–22 (in Japanese).
- Mayr E (1954) Changes in genetic environment and evolution. In: *Evolution as a Process* (eds Huxley JS, Hardy AC and Ford EB), pp. 157–180. Allen and Unwin, London.
- Mayr E (1963) *Animal Species and Evolution*. Harvard University Press, Cambridge, Massachusetts.
- Molecular Ecology Resources Primer Development Consortium *et al.* (2011) Permanent Genetic Resources added to Molecular Ecology Resources Database 1 August 2011–30 September 2011. *Molecular Ecology Resources*.
- Morgan K, Linton YM, Somboon P *et al.* (2010) Inter-specific gene flow dynamics during the Pleistocene-dated speciation of forest-dependent mosquitoes in Southeast Asia. *Molecular Ecology*, **19**, 2269–2285.
- Moritz C, Hoskin CJ, MacKenzie JB *et al.* (2009) Identification and dynamics of a cryptic suture zone in tropical rainforest. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 1235–1244.
- Moriyama EN, Gojobori T (1992) Rates of synonymous substitution and base composition of nuclear genes in *Drosophila*. *Genetics*, **130**, 855–864.
- Nadachowska K, Babik W (2009) Divergence in the face of gene flow: the case of two newts (Amphibia: Salamandridae). *Molecular Biology and Evolution*, **26**, 829–841.
- Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov Chain Monte Carlo approach. *Genetics*, **158**, 885–896.
- Niemiller ML, Fitzpatrick BM, Miller BT (2008) Recent divergence with gene flow in Tennessee cave salamanders (Plethodontidae: *Gyrinophilus*) inferred from gene genealogies. *Molecular Ecology*, **17**, 2258–2275.
- Nixon KC (1999) The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics*, **15**, 407–414.
- Nosil P (2008) Speciation with gene flow could be common. *Molecular Ecology*, **17**, 2103–2106.
- Osada N, Wu CI (2005) Inferring the mode of speciation from genomic data: a study of the great apes. *Genetics*, **169**, 259–264.
- Ota H (1998) Geographic patterns of endemism and speciation in amphibians and reptiles of the Ryukyu Archipelago, Japan, with special reference to their paleogeographical implications. *Researches on Population Ecology*, **40**, 189–204.
- Papadopoulou A, Anastasiou I, Vogler AP (2010) Revisiting the insect mitochondrial molecular clock: the mid-aegean trench calibration. *Molecular Biology and Evolution*, **27**, 1659–1672.
- Peters JL, Zhuravlev Y, Fefelov I, Logie A, Omland KE (2007) Nuclear loci and coalescent methods support ancient hybridization as cause of mitochondrial paraphyly between gadwall and falcated duck (*anas spp.*). *Evolution*, **61**, 1992–2006.
- Pinho C, Hey J (2010) Divergence with gene flow: models and data. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 215–230.
- Pinot S, Ramstein G, Harrison SP *et al.* (1999) Tropical paleoclimates at the last glacial maximum: comparison of Paleoclimate Modeling Intercomparison Project (PMIP) simulations and paleodata. *Climate Dynamics*, **15**, 857–874.
- Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253–1256.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Rull V (2008) Speciation timing and neotropical biodiversity: the Tertiary-Quaternary debate in the light of molecular phylogenetic evidence. *Molecular Ecology*, **17**, 2722–2729.
- Rull V (2011) Neotropical biodiversity: timing and potential drivers. *Trends in Ecology and Evolution*, **26**, 507–513.

- Sikes DS, Lewis PO (2001) *PAUPRat: PAUP* Implementation of the Parsimony Ratchet*. Distributed by the authors. Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut.
- Sokal RR, Crovello TJ (1970) The biological species concept: a critical evaluation. *American Naturalist*, **104**, 127–153.
- Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.
- Stephens M, Smith N, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.
- Strasburg JL, Rieseberg LH (2010) How robust are “isolation with migration” analyses to violations of the IM model? A simulation study. *Molecular Biology and Evolution*, **27**, 297–310.
- Swofford DL (2002) *PAUP* v. 4.0b10: Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer Associates, Sunderland, Massachusetts.
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.
- Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.
- Tamura K, Peterson D, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.
- Thalmann O, Fischer A, Lankester F, Paabo S, Vigilant L (2007) The complex evolutionary history of gorillas: insights from genomic data. *Molecular Biology and Evolution*, **24**, 146–158.
- Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, **3**, e285.
- Voris HK (2000) Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography*, **27**, 1153–1167.
- Wachowiak W, Palmé AE, Savolainen O (2011) Speciation history of three closely related pines *Pinus mugo* (T.), *P. uliginosa* (N.) and *P. sylvestris* (L.). *Molecular Ecology*, **20**, 1729–1743.
- Weekers PHH, De Jonckheere JF, Dumont HJ (2001) Phylogenetic relationships inferred from ribosomal ITS sequences and biogeographic patterns in representatives of the genus *Calopteryx* (Insecta: Odonata) of the west Mediterranean and adjacent west European zone. *Molecular Phylogenetics and Evolution*, **20**, 89–99.
- Wilgenbusch JC, Warren DL, Swofford DL (2004) AWTY: a system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference. Available from <http://ceb.csit.fsu.edu/awty>.
- Woerner AE, Cox MP, Hammer MF (2007) Recombination-filtered genomic datasets by information maximization. *Bioinformatics*, **23**, 1851–1853.
- Won YJ, Hey J (2005) Divergence population genetics of chimpanzees. *Molecular Biology and Evolution*, **22**, 297–307.
- Wu CI (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.
- Yeung CKL, Tsai PW, Chesser RT *et al.* (2011) Testing founder effect speciation: divergence population genetics of the spoonbills *Platalea regia* and *P. minor* (Threskiornithidae, Aves). *Molecular Biology and Evolution*, **28**, 473–482.
-
- Y.-H.L. is a master student at Tunghai University interested in the roles of islands and gene flow in speciation. C.-P.L. is a professor at Tunghai University with research interests in molecular systematics, character evolution and speciation of insects.
-

Data accessibility

DNA sequences: GenBank Accession nos EU603527, EU603531, EU603535, EU603537–EU603541, EU603556, EU603558–EU603559, EU603561–EU603563, EU603620, EU603631, EU603633, EU603635, EU603639–EU603640, EU603642, EU603644–EU603645, EU603649, EU603666, JF918934–JF918938, JF918940–JF918941, JF918944, JF918950–JF918956, JF918959, JN246927–JN247002, JN389796–JN390424, JN793699–JN793731, JN793751–JN793835, JQ974312–JQ974374, JQ995777–JQ995779.

Phylogenetic data: TreeBASE Study Accession no. S12507.

IMa2 input files: DRYAD entry doi: 10.5061/dryad.68s5c1r4.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1. Localities and collecting information for the *Euphaea* and outgroup specimens.

Table S2. Estimates of net genetic divergence per site and population migration rates for each of the 12 loci between *Euphaea* species.

Table S3. The number and mean time of inferred migration events for each locus between *Euphaea* species.

Table S4. Tests of nested demographic models for the two *Euphaea* species pairs based on the datasets of no recombined blocks, no non-neutral loci, and no recombined blocks and non-neutral loci.

Table S5. Maximum-likelihood estimates (MLE) and the 95% highest posterior density (HPD) intervals of demographic parameters estimated in IMA2 using the datasets of no recombined blocks (*E. formosa* & *E. yayeyamana*, 7780 bp; *E. decorata* and *E. ornata* 7921 bp), no non-neutral loci (*E. formosa* & *E. yayeyamana*, 6645 bp; *E. decorata* & *E. ornata*, 8396 bp), and no recombined blocks and non-neutral loci (*E. formosa* & *E. yayeyamana*, 5092 bp; *E. decorata* & *E. ornata*, 6774 bp).

Fig. S1 Phylogenies of *Euphaea* species derived from (A) maximum parsimony (MP), (B) maximum likelihood (ML) and (C) Bayesian inference (BI) analyses using multiple individuals.

Fig. S2 The log likelihood of the data for each (A) and for all individuals (B) of the four *Euphaea* species given a particular number of genetic clusters (*K*) from 1 to 6.

Fig. S3 The marginal posterior probability density distributions of model parameters (scaled by the mutation rate) estimated using the dataset of excluding recombined blocks in IMA2.

Fig. S4 The marginal posterior probability density distributions of model parameters (scaled by the mutation rate) estimated using the dataset of excluding non-neutral loci in IMA2.

Fig. S5 The marginal posterior probability density distributions of model parameters (scaled by the mutation rate) estimated

using the dataset of excluding both recombined blocks and non-neutral loci in IMA2.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.