

RESEARCH

Biogeography of *Drosophila* (Diptera: Drosophilidae) in East and Southeast AsiaFu-Guo Robert Liu,^{1,2} Shun-Chern Tsaur,^{3,*} and Hsiao-Ting Huang^{1,4,*}¹Department of Life Sciences, National Central University, Taoyuan 32001, Taiwan²Corresponding author, e-mail: liur1019@hotmail.com, liur@cc.ncu.edu.tw.³Division of Preparatory Programs for Overseas Chinese Students, National Taiwan Normal University, Linkou, New Taipei City 24449, Taiwan⁴Department of Life Sciences and Institute of Genome Sciences, National Yang-Ming University, Taipei 11221, Taiwan

*These authors contributed equally to this work.

Subject Editor: Johanne Brunet

J. Insect Sci. (2015) 15(1): 69; DOI: 10.1093/jisesa/iev056

ABSTRACT. The causes of high biological diversity in biodiversity hotspots have long been a major subject of study in conservation biology. To investigate this matter, we conducted a phylogeographic study of five *Drosophila* (Diptera: Drosophilidae) species from East and Southeast Asia: *Drosophila albomicans* Duda, *D. formosana* Duda, *D. immigrans* Sturtevant, *D. melanogaster* Meigen, and *D. simulans* Sturtevant. We collected 185 samples from 28 localities in eight countries. From each collected individual, we sequenced the autosomal extra sex comb gene (*esc*) and seven mitochondrial genes, including nicotinamide adenine dinucleotide hydrate-reductase dehydrogenase subunit 4 (*ND4*), *ND4L*, *tRNA-His*, *tRNA-Pro*, *tRNA-Thr*, partial *ND5*, and partial *ND6*. Phylogenetic analyses using maximum-likelihood and Bayesian methods revealed interesting population structure and identified the existence of two distinct *D. formosana* lineages (Southeast Asian and Taiwanese populations). Genetic differentiation among groups of *D. immigrans* suggests the possibility of endemic speciation in Taiwan. In contrast, *D. melanogaster* remained one extensively large population throughout East and Southeast Asia, including nearby islets. A molecular clock was used to estimate divergence times, which were compared with past geographical events to infer evolutionary scenarios. Our findings suggest that interglacial periods may have caused population isolation, thus enhancing population differentiation more strongly for some of the *Drosophila* species. The population structure of each *Drosophila* species in East and Southeast Asia has been influenced by past geographic events.

Key Words: differentiation, extra sex comb, inter-glacial, mitochondrion, phylogeography

Biologists have long been interested in the causes of high biodiversity (Schnitzler et al. 2011), and increasing evidence suggests that identifying biogeographic patterns can facilitate conservation efforts in biodiversity hotspot regions (Moritz and Faith 1998, Calsbeek et al. 2003, Mittermeier et al. 2004, Liu et al. 2006). Southeast Asia represents only 3% of the world's surface but is home to 20% of all known plant, animal, and marine species. Because of this richness, the region encompasses four of the world's 34 biodiversity hotspots (Lamoreux et al. 2006, Fig. 1). Many islands also exhibit high biodiversity, including 15 of the 34 hotspots (Lamoreux et al. 2006). Although not considered a hotspot, Taiwan has an endemic plants to area ratio of 2.96% (1,067/36,000), which is greater than that of the Caribbean (2.66%; 7,000/263,500), the fifth on the hotspot list (Mittermeier et al. 2004). Additionally, over 60% of the described insect species are endemic to Taiwan (Shao et al. 2003).

The Festoon Islands—including Japan, the Ryukyu Islands, Taiwan, and Luzon—run from north to south along the east coast of East and Southeast Asia. The present topography of these islands was formed by collisions between the Pacific Ocean, Philippine Sea plate and the Eurasian plate approximately 4–5 million years ago (MYA) (Faure et al. 1988, Huang et al. 2006). The tectonic movements and subsequent orogenesis created a rich diversity of geographic environments in Taiwan. Moreover, the Tropic of Cancer runs across central Taiwan, dividing the island into a tropical region in the south and a subtropical region in the north (Chen and Chen 2003, Fig. 1). Overall, Taiwan's geography provides an opportunity to investigate the population structure of taxa across diverse geographic landscapes (Turner et al. 2001).

For a century, *Drosophila* has been used as a model organism for studying behavior, development, genetics, gene regulation, and neurobiology. It is a favorite model organism because it has a short

development cycle and is easy and cheap to culture in laboratories, and a wide array of genetic tools are available for *Drosophila* studies. Although more than 3,000 *Drosophila* species have been recorded (Remsen and O'Grady 2002), most studies have focused on *Drosophila melanogaster* and a few other species. Therefore, population genetics data are lacking for most species of *Drosophila*. Fortunately, for many species, the phylogeny has been reconstructed using morphological and molecular data, providing basic information for future studies (Remsen and O'Grady 2002). The genus *Drosophila* consists of 10 subgenera, of which the two largest (*Drosophila* and *Sophophora*) comprise approximately 95% of the species within the genus (Markow and O'Grady 2005).

In this study, we selected five species (*D. albomicans* Duda, *D. formosana* Duda, *D. immigrans* Sturtevant, *D. melanogaster* Meigen, and *D. simulans* Sturtevant) from the two main subgenera to ensure ease of identification, comprehensive intra- and interspecies comparisons, and the inclusion of both closely and distantly related sympatric species. The first two species are distributed from East Asia to Southeast Asia, whereas the remaining three species have a cosmopolitan distribution (Markow and O'Grady 2005). We also hoped to see how human activities might impact their population genetic structure. Since the inclusion of multiple genes can facilitate the study of molecular evolution, here we used multiple genes—including a nuclear sequence (extra sex comb) and a mitochondrial DNA region encompassing four coding genes of nicotinamide adenine dinucleotide hydrate-reductase dehydrogenase (ND) and three tRNAs. In general, mitochondrial genes evolve approximately twice as fast as nuclear genes (DeSalle et al. 1987, Tamura 1992, Moriyama and Powell 1996). Thus, the subunits of ND—which are less conserved in vertebrates and invertebrates—are suitable for studying closely related species (Ferris et al. 1983, Moriyama and Powell 1997), whereas conserved nuclear genes, such as

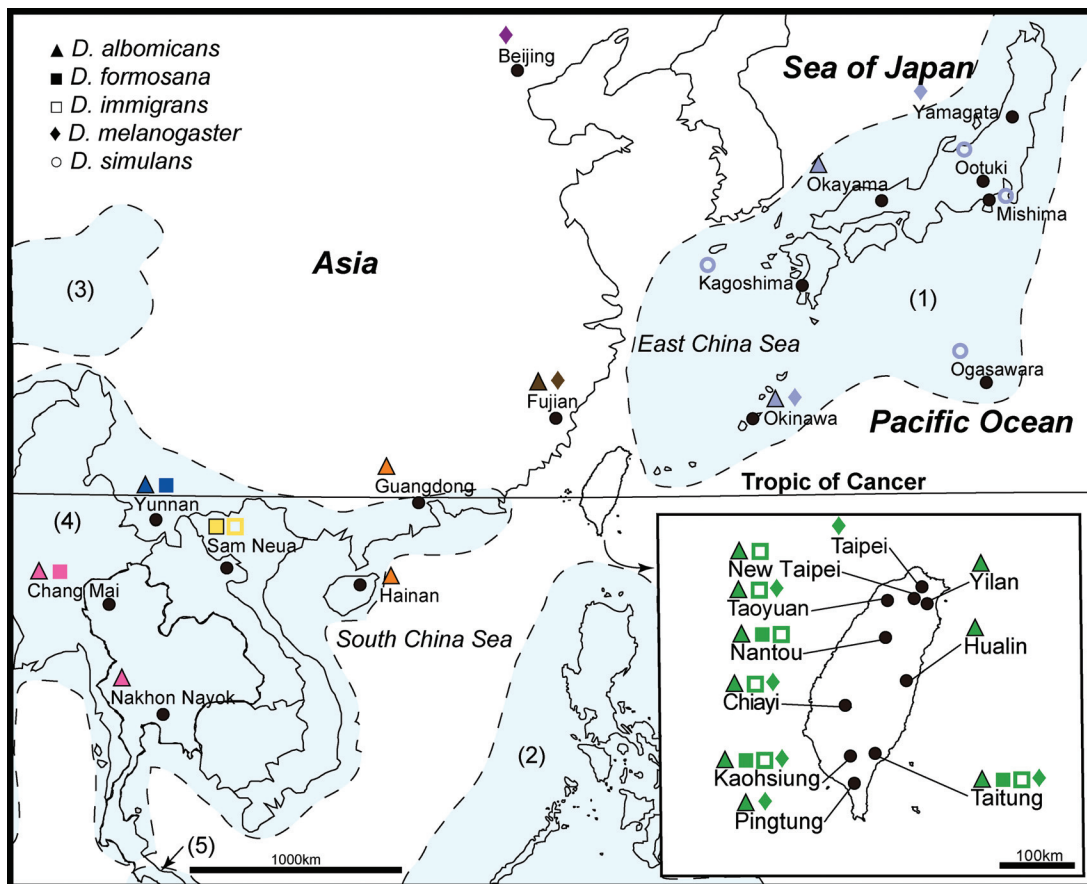


Fig. 1. Map of sampling localities. Solid circles indicate the sample localities. To reduce the map complexity, sample sizes of species collected from each locality are listed in [Supp Table 1](#) [online only]. The symbols at the top left side indicate the different species that were collected and labeled at each sample locality. The blue areas indicate the biodiversity hotspots in East and Southeast Asia: 1, Japan; 2, the Philippines; 3, the mountains of southwest China; 4, Indo-Burma; and 5, Sundaland (Lamoreux et al. 2006).

extra sex comb (*esc*), are appropriate for studying more distantly related taxa.

In this study, we were particularly interested in what population genetic structure and phylogeographic patterns would reveal on the biogeographic patterns in Southeast Asia. We examined whether different partitioning of the data (e.g., nuclear vs. mtDNA) provided similar topologies when reconstructing gene trees. Determining the date of divergence of species or populations enabled us to understand which geological events had led to high biodiversity. Here, we examined and compared the DNA sequence within and between species that shared the same geological history. Our results illustrate how geological events can shape phylogeographic patterns and affect the genetic differentiation of fruit flies in East and Southeast Asia.

Materials and Methods

Drosophila Samples. In total, 185 adult individuals were collected from 28 localities (Fig. 1; [Supp Table 1](#) [online only]). Upon collection, the males were immersed in 70% ethanol, while the females were brought back to the laboratory to lay eggs. Some samples were received from the Fly Stock Center of the Department of Biology, Ehime University, Ehime, Japan, and Drosophila Lab., Biodiversity Research Center, Academia Sinica, Taipei, Taiwan.

DNA Extraction. Each fly was placed in a 1.5-ml microcentrifuge tube containing 60 μ l protease solution (0.1 M Tris-HCl, pH 8.0; 0.05 M ethylenediaminetetraacetic acid; 0.2 M NaCl; 1% sodium dodecyl

sulfate; and 0.4 mg/ml protease K) (Beckenbach et al. 1993). The tubes were then incubated in a water bath at 65°C for 3 min to lyse the cell membranes. Phenol–chloroform was added to separate the proteins from the DNA. Next, 1 ml of 100% ethanol was added, and the mixture was maintained at –80°C for 30 min. The DNA was spun down at 13,000 rpm for 1 min, and the ethanol was decanted. Then the DNA was washed with 70% ethanol and spun at the same rate. After the ethanol was again decanted, the DNA was air dried and finally dissolved in 50–100 μ l ddH₂O and stored at –20°C for future experiments.

Polymerase Chain Reaction Amplification and DNA Sequencing.

Primers used to amplify the *esc* gene and a mitochondrial DNA fragment that included the complete *ND4*, *ND4L*, *tRNA-His*, *tRNA-Pro*, and *tRNA-Thr* genes and partial *ND5* and *ND6* genes were described in Kopp (2006) and Yu et al. (1999), respectively. For samples that were difficult to amplify, new mitochondrial DNA primers for the same region were designed using multiple species sequences from FlyBase (Tweedie et al. 2009). These new primer sequences were 5'-CCAGAACTGGAGCTTCAACATGAGC-3' and 5'-CGTTCTGGYTGATAWCCYCMHCCT-3'. The polymerase chain reaction (PCR) reaction mixture contained 1 \times PCR buffer, 1.5 mM of MgCl₂, 1 mM of dNTP, 1 M of primer, 5 ng of DNA sample, and 2.5 units of *Taq*. The thermal cycler was preheated to 94°C for 2 min; followed by 40 cycles of denaturation at 94°C for 50 s, annealing at 55°C for 2 min, and extension at 72°C for 1 min; with a final extension at 72°C for 10 min. The PCR products (3 μ l) were checked for size and quality by electrophoresis on a 1% agarose gel.

Then they were further purified using a PCR purification kit (One-Star Biotechnology Co., Ltd.). At the core facility at Academia Sinica, the samples were sequenced in both directions using a model 373 DNA Sequencer (35 cycles).

Data Analyses

Nucleotide and Amino Acid Sequence Comparison. Nucleotide sequences were aligned with MEGA 5 (Tamura et al. 2011). All single-nucleotide polymorphism sites in the original sequencing chromatograms were double checked for reading errors. Both aligned nucleotide and amino acid (AA) sequences were compared within and between sample groups.

Homogeneity Test. To test the efficacy of different types of sequences in reconstructing phylogeny, we performed partition homogeneity tests using PAUP* 4.0 (Swofford 2002) and the maximum likelihood ratio test (Huelsenbeck and Bull 1996) with a hundred simulation tests. PAUP* 4.0 (Swofford 2002) was used to calculate the likelihood values in the general time-reversible (GTR) +I (the proportion of invariable sites) +G (gamma distribution) model (Huelsenbeck and Bull 1996).

Based on the principle of “total evidence,” it has been suggested that individual gene trees or trees based on morphological characters cannot resolve many nodes of a phylogeny, whereas analyses of multiple genes can generate well-resolved phylogenies (Murrell et al. 2001). Moreover, parsimony evolutionary conflicts among nucleotide character states commonly occur in datasets, and the evolutionary congruence of nucleotide character states among datasets help to identify the most likely phylogeny (Kluge 1989, Rieppel 2009). Therefore, in cases of conflict between gene trees, we tested whether the different genes affected the robustness of the tree topology. In the absence of major conflicts between genes, the principle of “total evidence” was applied to our analyses to evaluate the relatedness between various populations and to estimate genetic differentiation, with the goal of understanding the history of these populations.

Genetic Variation and Sequence Divergence. Estimates of genetic diversity represent a valuable resource for biodiversity assessments (Goodall-Copestake et al. 2012). Haplotype (Berry et al. 1991) and nucleotide diversity (Nei 1978, Wang et al. 2002) represent the most commonly used measures of genetic diversity. Haplotype diversity (h) was calculated manually. Nucleotide diversity (π) and divergence were estimated using the maximum composite likelihood method in MEGA 5 (Tamura et al. 2011), which allows heterogeneous lineages and assumes a gamma distribution. Standard errors were obtained from 1,000 simulations. Comparisons of genetic divergence or nucleotide diversity among species were performed using analysis of variance (Sokal and Rohlf 2012).

Phylogenetic Analysis. A phylogenetic tree was reconstructed with Bayesian (Huelsenbeck et al. 2001) and maximum-likelihood analyses, using MrBayes 3.2 (Ronquist et al. 2012) and PhyML 3.0 (Guindon and Gascuel 2003) to elucidate the evolutionary relationships and to confirm consistency between the separated and combined gene datasets. Two runs of four chains each were run for 50,000,000 generations, with trees sampled every 100 generations. The first 10,000 trees were discarded as the burn-in. For both analyses, the best-fit substitution models—GTR+I+G and HKY+G—were selected using ModelTest v.3.7 (Posada and Crandall 1998, Posada and Buckley 2004) for each subgenera (*Drosophila* and *Sophophora*). For *Drosophila*, the parameter of gamma distribution shape (α) was 0.6323, and I was 0.5244. For *Sophophora*, α was 0.4492. We computed the bootstrap support (1,000 replicates) and posterior probabilities for the clades.

Gene Flow and F_{st} . Gene flow (Nm), or the genetic exchange between populations, was estimated using MIGRATE 3.5 (Beerli and Felsenstein 2001). The program was set at two replicates, each had one long chain with 10,000,000 generations and four short chains. The sample frequency was set at one out of every 100. The temperatures of the short chains were 1, 1.5, 3, and 10,000, as suggested in the menu. The burn-in was set to discard the first 20,000 trees. We monitored the analyses by establishing an acceptance ratio of greater than 30% and an

effective sample size of greater than 1,000, as well as by combining two separate runs to make parameter estimates. The F_{st} was calculated to examine population structure and the distribution of population differentiation when F_{st} was greater than 0.05 (Hartl and Clark 2007). The analyses were conducted using DnaSP v5.1 (Librado and Rozas 2009).

Haplotype Network. Haplotype network provides an alternative way to describe relationships between individuals and to clearly visualize substitution changes between sequences. The sequence data were used to estimate the genealogical relationships among the individuals. The inferred substitution steps represented the connections and differences between sequences. We used TCS 1.21 to perform this analysis (Clement et al. 2000). The cut-off of the parsimony probability was set at 95%.

Molecular Clock. Molecular clock uses AA changes during evolution to estimate the time of divergence and to match to historical events. The time of divergence was estimated using Bayesian analysis in a relaxed lognormal clock model with BEAST v1.7.5 (Drummond et al. 2012). We employed a Yule tree prior that assumed a constant lineage birth rate for each branch. The program was run for 100,000,000 generations, with a sampling frequency of one per 200. We referred to effective sample sizes of >200 and used the BEAST subprogram Tracer to evaluate the BEAST MCMC estimates and ensured that the analyses reached convergent status. The calibration time was set to 2.3 ± 0.65 million years (MY) for the divergence between *D. melanogaster* and *D. simulans* (Russo et al. 1995). The substitution rate of the GTR+G+I model and the parameters of prior probabilities remained at the default settings for the estimate.

Results

Sequence Data and Gene Homogeneity. For each individual, we sequenced a 2,184-bp mitochondrial (mt) fragment that included the full *ND4*, *ND4L*, *tRNA-His*, *tRNA-Pro*, and *tRNA-Thr* genes and partial *ND5* and *ND6* genes, as well as a 418-bp nuclear DNA fragment that included the partial *esc* gene. Among the 185 individuals from 28 localities, we found 143 and 37 distinct haplotypes for the mitochondrial and nuclear genes, respectively (accession numbers: JN418297–JN418476; Supp Table 1 [online only]). The alignments did not require the addition of any gap. Incomplete sequences were filled with “?” and these nucleotide positions (sites) were treated as missing data. Among the mitochondrial sequences, we identified 1,345 constant (61.6%), 839 variable (38.4%), 420 singleton (19.2%), and 419 parsimony informative (19.2%) sites. The nuclear fragments consisted of 292 constant (67.9%), 138 variable (32.1%), 29 singleton (6.9%), and 109 parsimony informative (26.1%) sites.

The gene homogeneity tests did not reveal major conflicts among the genes. Our use of Huelsenbeck and Bull’s (1996) maximum likelihood ratio test for homogeneity revealed no major conflicts among our genes with the GTR+I+G model ($P = 0.39$). Therefore, we used the combined gene dataset for further analyses.

Phylogeny and Biogeographic Units. The Bayesian and maximum-likelihood analyses yielded similar topologies. Except for *D. formosana* which was sorted into two strongly supported paraphyletic lineages (Fig. 2B), all other species were represented by monophyletic clades each with a strong bootstrap value ($\geq 89\%$) (Fig. 2; Supp Fig. 1 [online only]). Surprisingly, the *D. formosana* samples from the Asian mainland were more closely related to the *D. immigrans* samples than to *D. formosana* samples from Taiwan (Fig. 2B). The nucleotide divergence between these two *D. formosana* groups ($d = 0.061 \pm 0.008$) was significantly greater than that between the sister taxa *D. melanogaster* and *D. simulans* ($d = 0.051 \pm 0.009$; $P < 0.0001$; Table 1). The gene flow estimates between the two *D. formosana* lineages indicated low levels of gene flow ($Nm = 0.0002\text{--}0.0004$, estimated from F_{st} values). At the 95% connection limit, the haplotype network also suggested two distinct groups of *D. formosana* (Fig. 3)—Asia continental (DfCont) and Taiwanese (DfTW)—which are further discussed below.

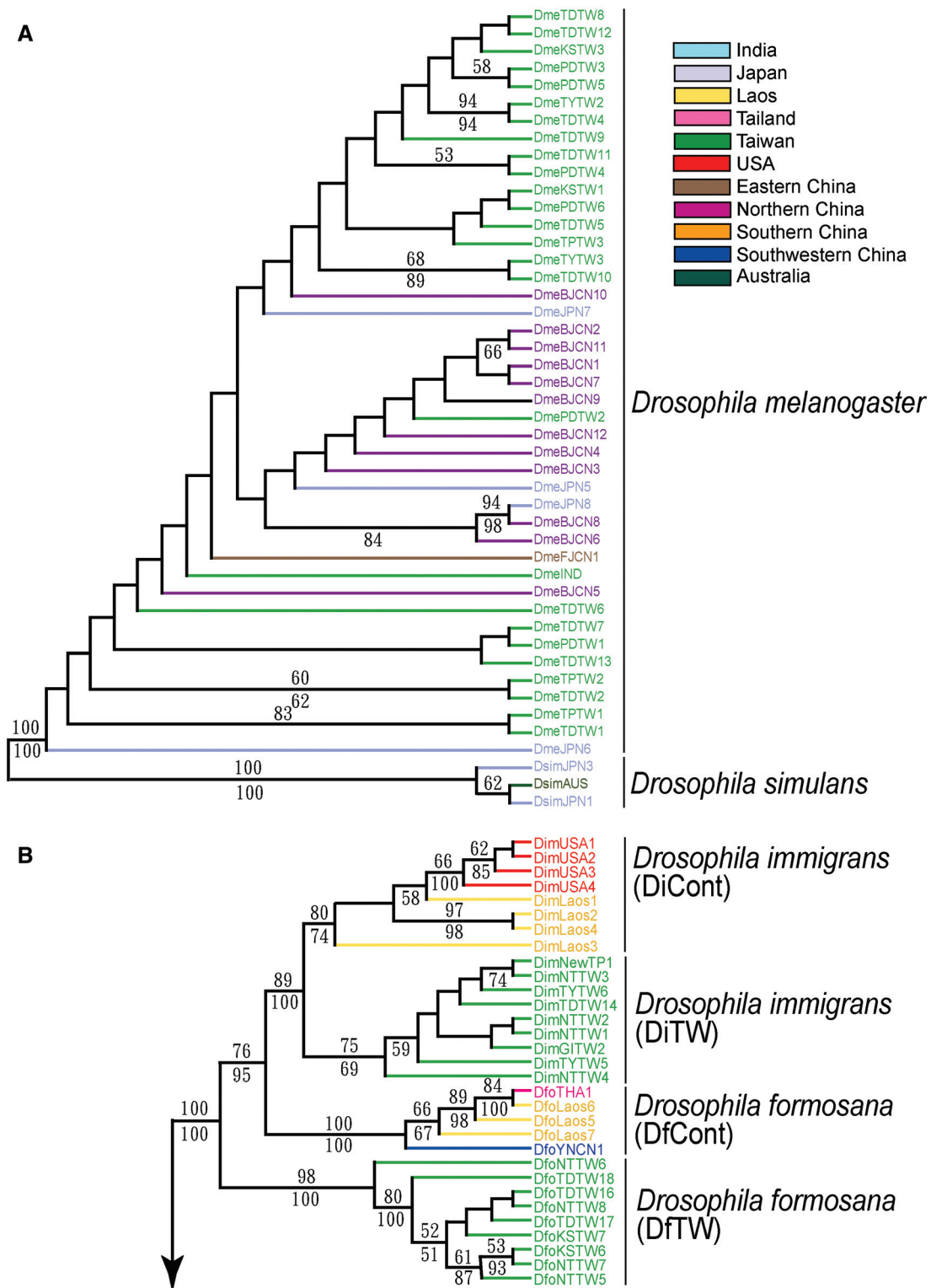


Fig. 2. Phylogenetic cladogram of five species of the genus *Drosophila*. Phylogenetic analyses were based on combined nuclear (*esc*) and mtDNA (*ND4*, *ND4L*, *ND5*, and *ND6*) genes to create separate estimates for the subgenera *Sophophora* (A) and *Drosophila* (B, C). The tree estimated using maximum likelihood analysis was adopted as the representative because it is conservative and similar to that of the Bayesian methods. The numbers above and below the branches denote the bootstrap support and posterior probabilities, respectively. Only support $\geq 50\%$ are shown. Taxonomic groups are indicated with vertical bars on the right-hand side. The arrows indicate the connections between two parts of the tree (B, C). DfCont and DiCont indicate sample clusters from the Asian continent, and DfTW and DiTW represent the sample clusters from Taiwan.

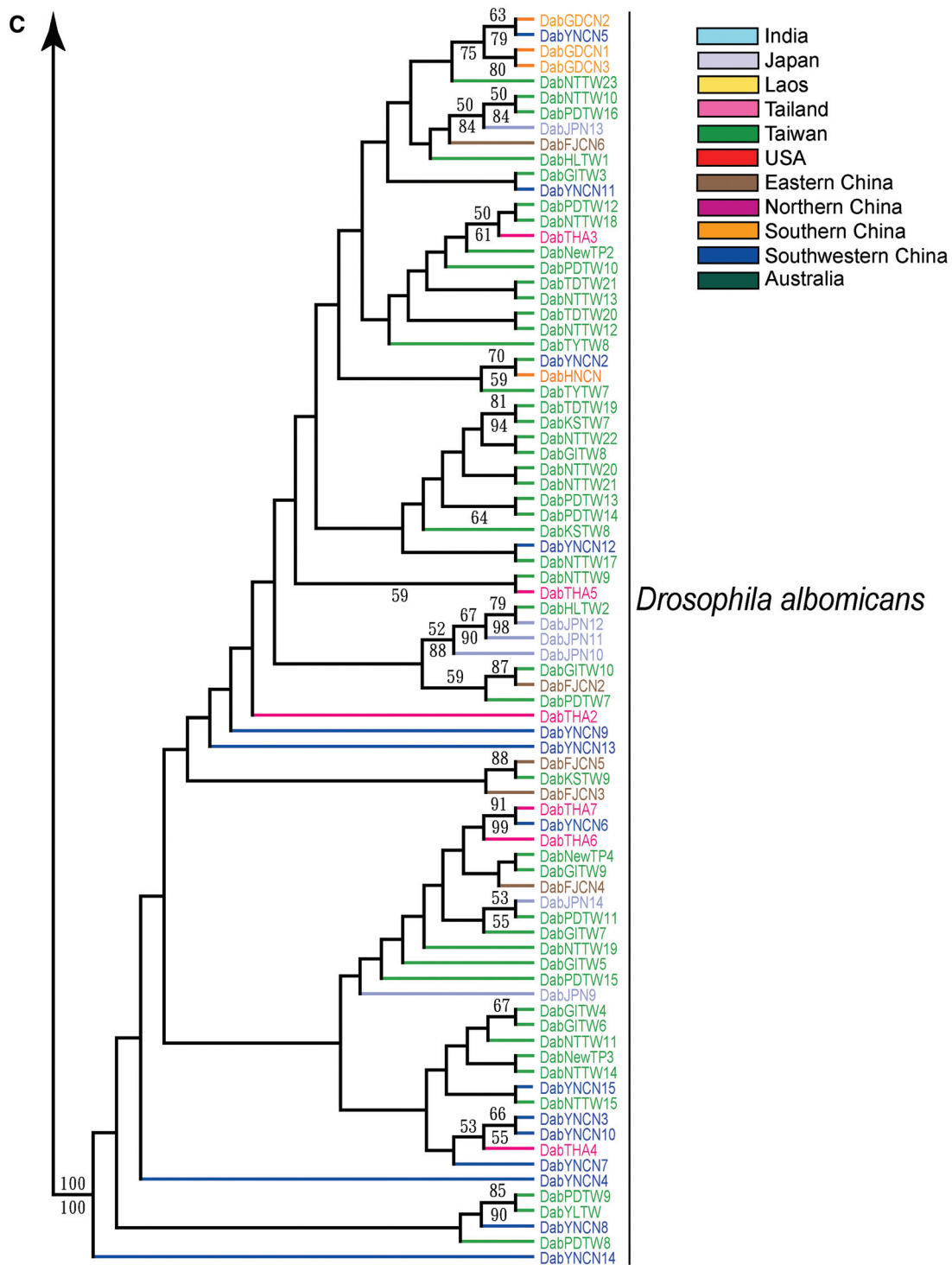


Fig. 2. Continued

In addition, the *D. immigrans* samples separated into two strongly supported (74–80% bootstrap) monophyletic subgroups (Fig. 2B). In the haplotype network, these two subgroups were 13 steps apart (Fig. 3) with low gene flow ($Nm = 0.0008\text{--}0.0007$). One group consisted of the Taiwanese samples (DiTW), while the other grouped samples from Laos and the United States (DiCont) (Fig. 2B).

In contrast, *D. albomicans* and *D. melanogaster* each formed a large monophyletic group with several weakly supported subgroups (Fig. 2A and C; Supp Fig. 1 [online only]). The haplotype network for each of these species was highly interconnected (Supp Fig. 2 [online only]).

We examined pairwise F_{st} values between some of the subgroups and obtained significant level of population differentiation within both *D. albomicans* (F_{st} : 0.085–0.126) and *D. melanogaster* (F_{st} : 0.116–0.143). However, groupings were unresolved due to low phylogenetic support within the *D. albomicans* and *D. melanogaster* populations.

Intra- and Interspecific Genetic Variation. At the intraspecific level, most species had small nucleotide diversity ($\pi < 0.014$; Table 2) approximately half that of the net interspecific divergence ($d \geq 0.031$, Table 1), except *D. formosana* ($\pi = 0.032$; Table 2). Separate analyses of the *D. formosana* subgroups led to considerably lower nucleotide

Table 1. Net nucleotide divergence (d) between taxa in five species of *Drosophila* in East and Southeast Asia

	<i>D. albomicans</i>	<i>D. formosana</i>	DfCont	DfTW	<i>D. immigrans</i>	DiCont	DiTW	<i>D. melanogaster</i>
<i>D. formosana</i>	0.147 (0.021)							
DfCont	0.141 (0.019)	NA						
DfTW	0.150 (0.022)	NA	0.061 (0.008)					
<i>D. immigrans</i>	0.134 (0.019)	0.046 (0.006)	0.033 (0.005)	0.052 (0.007)				
DiCont	0.134 (0.019)	0.033 (0.006)	0.035 (0.005)	0.055 (0.007)	NA			
DiTW	0.134 (0.018)	0.035 (0.006)	0.031 (0.004)	0.049 (0.007)	NA	0.011 (0.003)		
<i>D. melanogaster</i>	0.225 (0.035)	0.190 (0.029)	0.190 (0.028)	0.193 (0.029)	0.183 (0.027)	0.183 (0.027)	0.182 (0.027)	
<i>D. simulans</i>	0.173 (0.026)	0.156 (0.023)	0.156 (0.021)	0.170 (0.023)	0.158 (0.022)	0.162 (0.022)	0.153 (0.021)	0.051 (0.009)

The estimates were obtained using the maximum composite likelihood method in MEGA 5, allowing heterogeneous lineages and gamma distribution. The numbers in parentheses denote standard errors, which were obtained from 1,000 simulations. NA, not applicable. DfCont and DiCont refer to the sample clusters from the Asian continent, and DfTW and DiTW refer to the sample clusters from Taiwan.

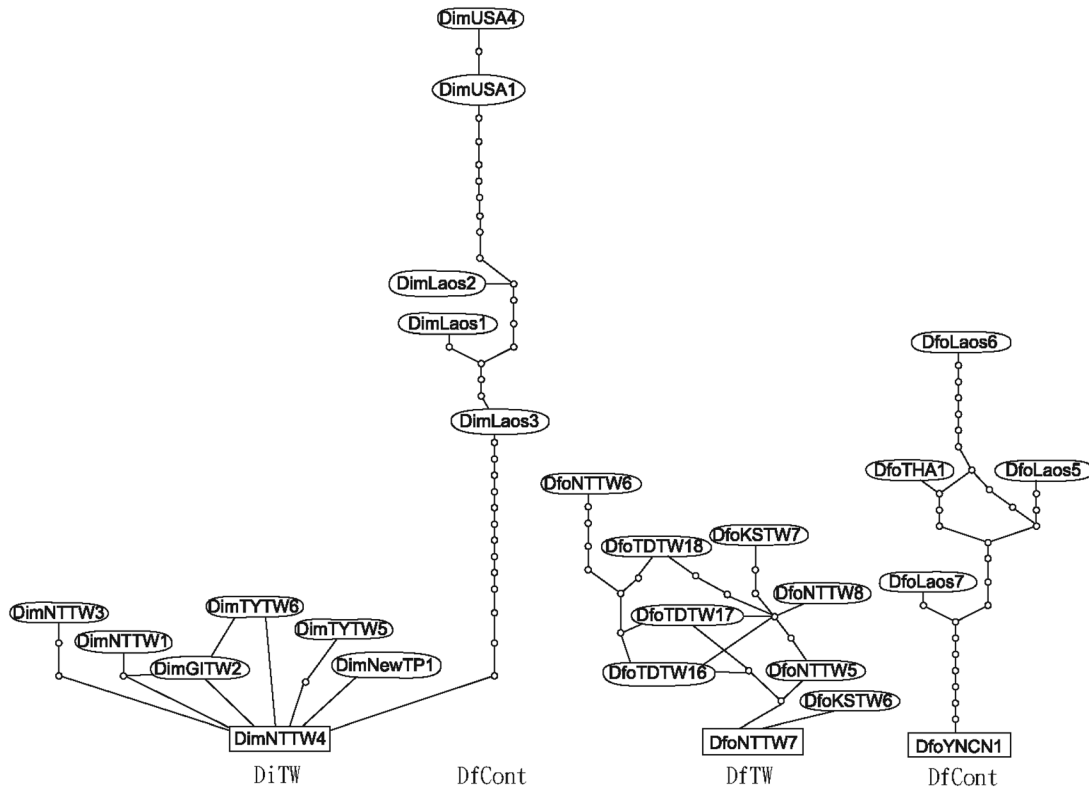


Fig. 3. The haplotype network of *D. formosana* and *D. immigrans*. The maximum number of mutational connections between pairs of sequences was set to a limit of 95%. Each internode circle represents a nucleotide change. When the probability of a parsimony connection was <95%, the samples were left alone with no linkage.

Table 2. Genetic variation among species groups

Species group	Sample size	Haplotype diversity	Nucleotide diversity (π)
<i>D. albomicans</i>	86	0.94	0.0138 (0.003)
<i>D. formosana</i>	15	1	0.0321 (0.004)
DfCont	5	1	0.0065 (0.002)
DfTW	10	1	0.0056 (0.002)
<i>D. immigrans</i>	22	0.91	0.0076 (0.002)
DiCont	12	0.92	0.0052 (0.002)
DiTW	10	0.9	0.0017 (0.001)
<i>D. melanogaster</i>	61	0.85	0.0082 (0.002)
<i>D. simulans</i>	5	0.6	0.0008 (0.001)

The numbers in parentheses are the standard errors. DfCont and DiCont refer to the sample clusters from the Asian continent, and DfTW and DiTW refer to the sample clusters from Taiwan.

diversity of 0.007 for DfCont and 0.006 for DfTW ($P < 0.001$; Table 2; Supp Fig. 1B [online only]). The highest nucleotide diversity was observed in *D. albomicans* ($\pi = 0.0138$).

At the interspecific level, our data suggested two strongly differentiated groups under *D. formosana* (Table 1). The net divergence between DfCont and DfTW ($d = 0.061 \pm 0.008$) was significantly greater than that between the well-known sister taxa *D. melanogaster* and *D. simulans* ($d = 0.051 \pm 0.009$) ($P < 0.0001$; Table 1; Fig. 2; Supp Fig. 1 [online only]). An F_{st} of 0.893 also indicated strong differentiation between the two *D. formosana* groups, DfCont and DfTW. Furthermore, the net divergence ($d = 0.033 \pm 0.005$) between DfCont and *D. immigrans* (Table 1) was approximately half of that between DfCont and DfTW. In contrast, DfTW was more distinct from *D. immigrans* (Table 1). AA changes between *D. formosana* and *D. immigrans*

did not reveal any trace of unique fixed AAs shared between DfCont and DfTW (Supp Fig. 3 [online only]). Of the 17 AA differences, five supported the DfCont and *D. immigrans* grouping, three supported the DfTW and *D. immigrans* grouping, but none supported the DfCont and DfTW grouping. Among the remaining nine AA differences, six polymorphisms were singletons, and the other three were not consistent with any taxonomic group. At the nucleotide level, the aforementioned clusters were supported by 44, 29, and 8 parsimony informative sites, respectively (Supp Fig. 4 [online only]). Therefore, the data support strong genetic differences between DfCont and DfTW.

Discussion

The Biogeography of *D. formosana*. Our results strongly support the presence of two genetically distinct groups of *D. formosana*. The two groups, DfCont and DfTW, formed well-resolved clades (bootstrap support $\geq 98\%$), showed strong levels of genetic differentiation ($d = 0.061 \pm 0.008$), had no shared fixed AAs, and formed independent groups in the haplotype networks. Furthermore, within a 2-hour observation, no physical mating between DfCont and DfTW has been found under controlled experiments (Ting 1997), and no hybrids have been discovered in reciprocal cross tests or in the wild where both species were collected together (C. T. Ting, personal communication). Taken together, this evidence strongly supports the notion that these two genetically populations of *D. formosana* may have evolved into two distinct species. Therefore, one of these populations should be considered a cryptic species. Since the type locality of *D. formosana* is Taiwan (Duda 1923), the DfTW in Taiwan may represent the original *D. formosana*. Moreover, DfTW is most distinct from *D. immigrans*, which may also suggest a more ancient lineage. Thus, the samples collected from other Southeast Asia locales—including Xishuangbanna, Chiangmai, Borneo, and Indonesia—require a new name. However, sample size was limited in this study, and more regions and individuals must be sampled before our finding can be strongly confirmed. It is indeed possible that additional sampling would reveal additional geographical structure or even potentially some degree of genetic exchange among clades although the lack of offspring and hybrid production between the two groups may limit this possibility. Our results do at the very least suggest interesting directions for future studies.

Based on Bayesian estimates, DfCont and DfTW were separated about 3.54 ± 0.14 MYA. Geological studies have shown that during the mid-Pliocene era (approximately 3–3.3 MYA), the average temperature was 2–3°C higher than today (Robinson et al. 2008), and the sea level was close to 25 m above the current sea level (Müller et al. 2008, Dwyer and Chandler 2009). The ocean was a natural barrier isolating islands from the mainland and enhancing the independent evolution of isolated populations.

The Biogeography of *D. immigrans*. The *D. immigrans* population in Taiwan (DiTW) formed a separate clade from the populations from Laos and the United States, and the two groups were separated by 13 mutations in the haplotype network. In addition, the two groups were strongly differentiated genetically, F_{st} of 0.521, and their DNA sequences contained nine nucleotides fixed in each clade. A hypothesis may be suggested that ancestors to DiTW may have been separated from DiCont approximately 2.11 ± 0.05 MYA, and their populations could have been repeatedly separated during periods of interglacial high sea level in the Pleistocene (approximately 0.117–2.588 MYA) (Richmond and Fullerton 1986, Müller et al. 2008, Ruen 2011). Future sampling of this species over its range is necessary before we can confirm whether *D. immigrans* in Taiwan merits a qualification as a subspecies.

The Biogeography of *D. albomicans* and *D. melanogaster*. The samples of *D. albomicans* and *D. melanogaster* presented two large populations with potential population structure in East and Southeast

Asia. Further investigation is required to assess the shallow population structure, which may result from limited sampling but could also illustrate the impact of movement of these species by humans. Within *D. melanogaster*, samples from the same locality tended to cluster together, while, in contrast, several clades within *D. albomicans* included geographically distant samples. These patterns could suggest differences between these two species where populations of *D. melanogaster* in East and Southeast Asia are relatively young, while the current *D. albomicans* population is derived from multiple lineages dating back to their common ancestor of around 2.58 ± 0.07 MYA (Gibbard and Cohen 2007).

The Absence of *D. simulans* in Taiwan. *D. simulans* has not yet been discovered in Taiwan despite the fact that it has been recognized as a cosmopolitan species and is distributed on the Asian mainland. Furthermore, it is found throughout the Ryukyu Islands, which are much smaller than Taiwan and further from the Asian mainland and in the Philippines which is a great distance from the Asian mainland. The morphological characteristic that distinguishes *D. melanogaster* from *D. simulans* is the posterior lobe of the male genital arch. We checked all collected males and the male offspring from the isofemale lines caught in the wild. Yet, no fruit flies were identified as *D. simulans* as has been observed by other collectors (M. Toda, M. Kimura, H. Watabe, M. Watada, personal communication). This finding may result from seasonal population fluctuations as no *D. simulans* could be collected in China during summer (Torres and Madi-Ravazzi 2006). Future studies should sample for *D. simulans* throughout the year to determine whether it is present in Taiwan. However, the extremely low genetic differences observed between individuals from Australia and Japan indicate that there is one widely distributed population of *D. simulans* extending from East Asia to Australia.

Biogeographic Patterns. The species *D. formosana* and *D. immigrans* both included highly differentiated genetic groups, while *D. melanogaster* and *D. simulans* each consisted of one large genetic population in East and Southeast Asia. Our data indicate similarity in genetic structure between closely related sympatric species but differences among distant taxa.

The species *D. formosana*, and *D. immigrans* were apparently more isolated during interglacial periods when temperatures were warmer and sea levels higher. Interglacial periods may have caused population isolation and enhanced population differentiation in other species.

Supplementary Data

Supplementary data are available at *Journal of Insect Science* online.

Acknowledgments

This work was supported by the National Science Council of Taiwan, ROC under Grant NSC97-2621-B-008-001. We are grateful to Chau-Ti Ting (National Taiwan University), Masami Hayashi (Saitama University), and Masayoshi Watada (Ehime University), for providing flies and assistance. Special thanks also go to the Fu-Shan Botanical Garden for the samples from Taiwan, and to Hongwei Chen (South China Agricultural University) and Wenxia Zhang (Peking University) for conducting fieldwork in China.

References Cited

- Beckenbach, A. T., Y. W. Wei, and H. Liu. 1993. Relationships in the *Drosophila obscura* species group, inferred from mitochondrial Cytochrome Oxidase II sequences. *Mol. Biol. Evol.* 10: 619–634.
- Beerli, P., and J. Felsenstein. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. *Proc. Natl. Acad. Sci. USA.* 98: 4563–4568.
- Berry, A. J., J. W. Ajioka, and M. Kreitman. 1991. Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. *Genetics* 129: 1111–1117.

- Calsbeek, R., J. N. Thompson, and J. E. Richardson. 2003.** Patterns of molecular evolution and diversification in a biodiversity hotspot: the California floristic province. *Mol. Ecol.* 12: 1021–1029.
- Chen C. S., and Y. L. Chen. 2003.** The rainfall characteristics of Taiwan. *Mon. Weather Rev.* 131: 1323–1341.
- Clement, M., D. Posada, and K. A. Crandall. 2000.** TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* 9: 1657–1659.
- DeSalle, R., T. Freedman, E. M. Prager, and A. C. Wilson. 1987.** Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila*. *J. Mol. Evol.* 26: 157–164.
- Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut. 2012.** Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29: 1969–1973.
- Duda, O. 1923.** Die orientalischen und australischen *Drosophiliden*-Arten (Dipteren) des ungarischen National-Museums Zu Budapest. *Annales Musei Nationalis Hungarici* 20: 241–250.
- Dwyer, G. S., and M. A. Chandler. 2009.** Mid-Pliocene sea level and continental ice volume based on coupled benthic Mg/Ca palaeotemperatures and oxygen isotopes. *Philos. Trans. A Math. Phys. Eng. Sci.* 367: 157–168.
- Faure, M., P. Monié, and O. Fabbri. 1988.** Microtectonics and 39Ar–40Ar dating of high pressure metamorphic rocks of the south Ryukyu Arc and their bearings on the pre-Eocene geodynamic evolution of Eastern Asia. *Tectonophysics* 156: 133–143.
- Ferris, S. D., R. D. Sage, C. M. Huang, J. T. Nielsen, U. Ritte, and A. C. Wilson. 1983.** Flow of mitochondrial DNA across a species boundary. *Proc. Natl. Acad. Sci. USA.* 80: 2290–2294.
- Gibbard, P. L., and K. M. Cohen. 2007.** Global chronostratigraphical correlation table for the last 2.7 million years. *Episodes* 31: 243–247.
- Goodall-Copestake, W. P., G. A. Tarling, and E. J. Murphy. 2012.** On the comparison of population-level estimates of haplotype and nucleotide diversity: a case study using the gene *cox1* in animals. *Heredity* 109:50–56.
- Guindon, S., and O. Gascuel. 2003.** A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52: 696–704.
- Hartl, D. L., and A. G. Clark. 2007.** Principles of population genetics. Sinauer Associates, Sunderland, MA.
- Huang, C. Y., P. B. Yuan, and S. J. Tsao. 2006.** Temporal and spatial records of active arc-continent collision in Taiwan: a synthesis. *Geol. Soc. Am. Bull.* 118: 274–288.
- Huelsenbeck, J. P., and J. J. Bull. 1996.** A likelihood ratio test to detect conflicting phylogenetic signal. *Syst. Biol.* 45: 92–98.
- Huelsenbeck, J. P., F. Ronquist, R. Nielsen, and J. P. Bollback. 2001.** Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294: 2310–2314.
- Kluge, A. G. 1989.** A concern for evidence, and a phylogenetic hypothesis of relationships amongst Epicrates (Boidae, Serpentes). *Syst. Zool.* 38: 7–25.
- Kopp, A. 2006.** Basal relationships in the *Drosophila melanogaster* species group. *Mol. Phylogenet. Evol.* 39: 787–798.
- Lamoreux, J. F., J. C. Morrison, T. H. Ricketts, D. M. Olson, E. Dinerstein, M. W. McKnight, and H. H. Shugart. 2006.** Global tests of biodiversity concordance and the importance of endemism. *Nature* 440: 212–214.
- Librado, P., and J. Rozas. 2009.** DnaSP v5.1: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
- Liu, F.-G. R., P. E. Moler, and M. M. Miyamoto. 2006.** Phylogeography of the salamander genus *Pseudobranchius* in the southeastern United States. *Mol. Phylogenet. Evol.* 39: 149–159.
- Markow, T. A., and P. M. O’Grady. 2005.** *Drosophila*: a guide to species identification and use. Academic Press, London.
- Mittermeier, R. A., P. R. Gil, M. Hoffmann, J. D. Pilgrim, T. M. Brooks, C. G. Mittermeier, J. F. Lamoreux, and G.A.B. da Fonseca. 2004.** Hotspots revisited: earth’s biologically richest and most endangered terrestrial ecoregions. University of Virginia Press, Sierra Madre, CA.
- Moritz, C., and D. P. Faith. 1998.** Comparative phylogeography and the identification of genetically divergent areas for conservation. *Mol. Ecol.* 7: 419–429.
- Moriyama, E. N., and J. R. Powell. 1996.** Intraspecific nuclear DNA variation in *Drosophila*. *Mol. Biol. Evol.* 13: 261–277.
- Moriyama, E. N., and J. R. Powell. 1997.** Synonymous substitution rates in *Drosophila*: mitochondrial versus nuclear genes. *J. Mol. Evol.* 45: 378–391.
- Müller, R. D., M. Sdrolas, C. Gaina, B. Steinberger, and C. Heine. 2008.** Long-term sea-level fluctuations driven by ocean basin dynamics. *Science* 319: 1357–1362.
- Murrell, A., N.J.H., Capbell, and S. C. Barker. 2001.** A total-evidence phylogeny of ticks provides insights into the evolution of life cycles and biogeography. *Mol. Phylogenet. Evol.* 21: 244–258.
- Nei, M. 1978.** Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583–590.
- Posada, D., and T. R. Buckley. 2004.** Model selection and model averaging in phylogenetics: advantages of the AIC and Bayesian approaches over likelihood ratio tests. *Syst. Biol.* 53: 793–808.
- Posada, D., and K. A. Crandall. 1998.** MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817–818.
- Remsen, J., and P. O’Grady. 2002.** Phylogeny of *Drosophilinae* (Diptera: Drosophilidae), with comments on combined analysis and character support. *Mol. Phylogenet. Evol.* 24: 249–264.
- Richmond, G. M., and D. S. Fullerton. 1986.** Summation of Quaternary glaciations in the United States of America. *Quaternary Sci. Rev.* 5: 183–196.
- Rieppel, O. 2009.** ‘Total evidence’ in phylogenetic systematics. *Biol. Philos.* 24: 607–622.
- Robinson, M., H. J. Dowsett, and M. A. Chandler. 2008.** Pliocene role in assessing future climate impacts. *EOS Trans. Am. Geophys. Union.* 89: 501–502.
- Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard, and J. P. Huelsenbeck. 2012.** MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61: 539–542.
- Ruen, T. 2011.** Ice core data estimates of Atmospheric CO₂ over the last 800k years. (http://en.wikipedia.org/wiki/File:Co2_glacial_cycles_800k.png).
- Russo, C.A.M., N. Takezaki, and M. Nei. 1995.** Molecular phylogeny and divergence times of drosophilid species. *Mol. Biol. Evol.* 12: 391–404.
- Schnitzler, J., T. G. Barraclough, J. S. Boatwright, P. Goldblatt, J. C. Manning, M. P. Powell, T. Rebelo, and V. Savolainen. 2011.** Causes of plant diversification in the cape biodiversity hotspot of South Africa. *Syst. Biol.* 60: 343–357.
- Shao, K. T., H. Lee, and Y. C. Lin. 2003.** TaiBNET. (http://taibnet.sinica.edu.tw/chi/taibnet_reference_sum.php).
- Sokal, R. R., and F. J. Rohlf. 2012.** Biometry, 4th ed. W. H. Freeman and Company, New York.
- Swofford, D. L. 2002.** PAUP*: phylogenetic analysis using parsimony (* and other methods). Sinauer Associates, Sunderland, MA.
- Tamura, K. 1992.** Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C-content biases. *Mol. Biol. Evol.* 9: 678–687.
- Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar. 2011.** MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731–2739.
- Ting, C.T. 1997.** Genetic analysis of population differentiation and premating isolation. PhD thesis, National Taiwan University, Taipei, Taiwan.
- Torres, F. R., and L. Madi-Ravazzi. 2006.** Seasonal variation in natural populations of *Drosophila* spp. (Diptera) in two woodlands in the State of São Paulo, Brazil. *Iheringia Série Zoologia* 96: 427–434.
- Turner, H., P. Hovenkamp, and P. C. van Welzen. 2001.** Biogeography of Southeast Asia and the west Pacific. *J. Biogeogr.* 28: 217–230.
- Tweedie, S., M. Ashburner, K. Falls, P. Leyland, P. Mcquilton, S. Marygold, G. Millburn, D. Osumi-Sutherland, A. Schroeder, R. Seal, et al. 2009.** Flybase: enhancing *Drosophila* gene ontology annotations. *Nucleic Acids Res.* 37: D555–D559.
- Wang, W., K. Thornton, A. Berry, and M. Long. 2002.** Nucleotide variation along the *Drosophila melanogaster* fourth chromosome. *Science* 295: 134–137.
- Yu, H., W. Wang, S. Fang, Y. P. Zhang, F. J. Lin, and Z. C. Geng. 1999.** Phylogeny and evolution of the *Drosophila nasuta* subgroup based on mitochondrial *ND4* and *ND4L* gene sequences. *Mol. Phylogenet. Evol.* 13: 556–565.

Received 9 February 2014; accepted 11 May 2015.