Molecular Phylogenetics and Evolution 60 (2011) 98–107

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Phylogenetic relationships between *Sophophora* and *Lordiphosa*, with proposition of a hypothesis on the vicariant divergences of tropical lineages between the Old and New Worlds in the family Drosophilidae

Jian-jun Gao [a,1], Yao-guang Hu [b,1], Masanori J. Toda [b,c,*], Toru Katoh [d], Koichiro Tamura [e]

[a] Laboratory for Conservation and Utilization of Bio-resources, Yunnan University, 2 North Green Lake Road, Kunming, Yunnan, China
[b] Institute of Low Temperature Science, Hokkaido University, N19 W8, Kita-ku, Sapporo, Japan
[c] The Hokkaido University Museum, N10 W8, Kita-ku, Sapporo, Japan
[d] Graduate School of Science, Hokkaido University, N10 W8, Kita-ku, Sapporo, Japan
[e] Department of Biological Sciences, Tokyo Metropolitan University, Tokyo, Japan

ARTICLE INFO

Article history:
Received 18 October 2010
Revised 11 April 2011
Accepted 18 April 2011
Available online 24 April 2011

Keywords:
Eocene
North Atlantic Land Bridge
Nucleotide compositional heterogeneity
Paraphyly
Phylogeography
Sequence partitioning

ABSTRACT

Despite many studies on the phylogeny of the subgenus *Sophophora*, its monophyly has not been established, especially in relation to its putative relative, the genus *Lordiphosa*. We analyzed their phylogenetic relationships using DNA sequence data of two mitochondrial genes (*ND2* and *COII*) and two nuclear genes (*Adh* and *28SrRNA*). In constructing phylogenetic trees, we accounted for the problem of among-taxa nucleotide compositional heterogeneity, and took a sequence-partitioning approach to allow multiple substitution models for nucleotide sequences that have evolved under different evolutionary processes, particularly developing a novel, sequence-partitioning procedure for Neighbor Joining (NJ) tree construction. Trees constructed by different methods showed an almost identical and strongly supported topology in which *Sophophora* was paraphyletic: *Lordiphosa* was placed as the sister to the Neotropical *Sophophora* consisting of the *saltans* and *willistoni* groups, and *Sophophora* was divided into the clade of *Lordiphosa* + Neotropical *Sophophora* and the clade of the *obscura* + *melanogaster* groups. Based on the estimated time, 45.9 Mya, of divergence between the Old World *Lordiphosa* and the Neotropical *Sophophora* and evidence from paleontology, paleo-geography and -climatology, we propose a hypothesis that this vicariant divergence should have occurred when the North Atlantic Land Bridge between Europe and North America broke in the middle Eocene Epoch.

1. Introduction

The family Drosophilidae has long been studied in evolutionary biology, and its significance has recently been further manifested by the *Drosophila* 12 Genomes Project (Drosophila 12 Genomes Consortium, 2007). The released database of whole genome sequences of 12 *Drosophila* species should promote genome-based comparative studies in various fields of biology across the family Drosophilidae. Any comparative studies necessitate the baseline information on phylogenetic relationships of concerned taxa, but the phylogeny of Drosophilidae has not fully been resolved yet.

Of the 12 *Drosophila* species whose whole genomes have been sequenced, nine species belong to the subgenus *Sophophora* Sturtevant. This subgenus has been and shall continue to be one of the most important groups of model organisms in evolutionary biology. Of the eight species groups placed in this subgenus, the *melanogaster* [including the *ananassae* and *montium* groups raised by Da Lage et al. (2007)], *obscura*, *saltans*, *willistoni*, *fima*, *setifemur*, *populi* and *dentissima* groups (Burla, 1954; Lemeunier and Ashburner, 1976; Mather, 1955; McEvey, 2009; Sturtevant, 1942; Tsacas, 1980), the first four have been most intensively studied, with several model species such as *Drosophila melanogaster* Meigen, 1830 very important in all fields of modern biology and *Drosophila pseudoobscura* Frolova, 1929 of which population genetics has been studied by Th. Dobzhansky and colleagues since the 1930's (see Dobzhansky and Powell, 1975a, for review). Two Neotropical species groups, the *saltans* and *willistoni* groups, have been good materials for studying speciation mechanisms, with incipient species at various stages of speciation process (see Dobzhansky and Powell (1975b) and Ehrman and Powell (1982), for review of studies on the *willistoni* group).

Throckmorton (1975) treated the "Sophophoran radiation" consisting of the subgenus *Sophophora* and the genera *Chymomyza*

* Corresponding author at: Institute of Low Temperature Science, Hokkaido University, N19 W8, Kita-ku, 060-0819 Sapporo, Japan. Fax: +81 11 706 7142.
E-mail addresses: gao-leyun@263.net (J.-j. Gao), hu339@hotmail.com (Y.-g. Hu), hutian@pop.lowtem.hokudai.ac.jp (M.J. Toda), tkatoh@mail.sci.hokudai.ac.jp (T. Katoh), ktamura@tmu.ac.jp (K. Tamura).
[1] These authors contributed equally to the present study.

1055-7903/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2011.04.012

Czerny and *Neotanygastrella* Duda as one of basal radiations of the subfamily Drosophilinae. Since then, the monophyly of *Sophophora* has been attested by a number of phylogenetic studies involving at least the above-mentioned four major species groups of this subgenus, based on evidence from various sources, e.g., larval hemolymph protein (LHP) immunoprecipitation (Beverley and Wilson, 1982), gene structure (Powell and DeSalle, 1995; Wojtas et al., 1992), nucleotide sequences of different genes (Kwiatowski and Ayala, 1999; Kwiatowski et al., 1994; O'Grady and Kidwell, 2002), mitochondrial genome sequences (O'Grady and DeSalle, 2008), a combined analysis with morphological characters and DNA sequences (Remsen and O'Grady, 2002), a supertree analysis of 117 published phylogenetic trees (van der Linde and Houle, 2008) and a supermatrix-based molecular phylogenetic analysis (van der Linde et al., 2010). However, some studies, such as Pélandakis et al. (1991) and Pélandakis and Solignac (1993) based on the large subunit of ribosomal RNA (*28S*) gene and Kwiatowski et al. (1997) based on the glycerol-3-phosphate dehydrogenase (*Gpdh*) gene, did not support the monophyly of the subgenus *Sophophora*, but placed the Neotropical *saltans–willistoni* clade outside a large clade comprising all other *Drosophila* species (including the *melanogaster* and *obscura* groups) and related genera, though bootstrap support for this relationship was not so high. A probable source of bias having made these results incongruent with most of other studies supporting the monophyly of the subgenus *Sophophora* is neglecting of nucleotide composition variation among taxa. Tarrio et al. (2001) have pointed out that there is a wide variation in the guanine plus cytosine (GC) content of genes, especially at the third codon position, among the members of the family Drosophilidae (see also Anderson et al., 1993; Rodriguez-Trelles et al., 1999a,b; Tatarenkov et al., 2001); the *melanogaster* and *obscura* groups have high GC contents, while the *willistoni* group and the genus *Chymomyza* have the low contents. And, Tarrio et al. (2001) have clearly shown that correct phylogenetic assessment cannot be achieved when the heterogeneous base composition is not accounted for by the substitution model to be applied for molecular tree construction. Similarly, Da Lage et al. (2007) suggested mal-effects of base compositional heterogeneity on phylogenetic reconstruction for the *Amyrel* gene, a paralogue of the α-amylase genes: although the unweighted Maximum Parsimony (MP) analysis inferred the paraphyly of the subgenus *Sophophora*, placing the two Neotropical species groups with exceptionally low GC contents outside the major drosophiline clade, the monophyly of *Sophophora* was recovered when the third codon positions were downweighted in MP analyses.

An unsettled issue with respect to the phylogeny of the subgenus *Sophophora* is its relationship to the genus *Lordiphosa* Basden (O'Grady and Kidwell, 2002), of which species are distributed in the Oriental and Palearctic Regions. Katoh et al. (2000) suggested, using the alcohol dehydrogenase (*Adh*) gene sequences, that *Sophophora* was paraphyletic in respect to *Lordiphosa*, specifically corroborating the sister relationship between *Lordiphosa* and the Neotropical *saltans–willistoni* clade by high bootstrap support (95%). They further suggested that the clade comprising *Lordiphosa* and the *saltans* and *willistoni* groups was placed outside the major drosophiline lineage including the *melanogaster–obscura* clade, although this topology was not strongly supported and should have been biased by no account of the compositional heterogeneity among sequences studied. On the other hand, Hu and Toda (2001) suggested, based on a cladistic analysis with morphological characters, that *Sophophora* (the *melanogaster–obscura* clade) formed a monophyletic group with *Lordiphosa*, although its bootstrap support was not so high and the *saltans* and *willistoni* groups were not included in the analysis.

In this study, we seek to determine the phylogenetic relationships among the four major species groups of *Sophophora* and the four species groups of *Lordiphosa* by expanding taxon sampling for these focal taxa. We employed DNA sequence data of two mitochondrial genes, the NADH dehydrogenase subunit 2 (*ND2*) and the cytochrome *c* oxidase II (*COII*), and two nuclear genes, *Adh* and *28S*, and took into account variations in nucleotide composition among the studied sequences to provide valuable information for reconstructing the phylogeny. Finally, based on the results, we propose a phylogeographic hypothesis on one of major patterns in the evolution of Drosophilinae, i.e., the disjunction of tropical fauna between the Old and New Worlds.

2. Materials and methods

2.1. Taxon sampling

A total of 27 species were selected as ingroup taxa from the subfamily Drosophilinae, with particular emphasis on the four major species groups of *Sophophora* and the four species groups, the *denticeps*, *fenestrarum*, *miki* and *nigricolor* groups, of *Lordiphosa* (Table 1). As outgroup taxa, two species were selected from the genera *Scaptodrosophila* Duda and *Chymomyza* that have proved to represent basal lineages of the Drosophilinae in most of the previous molecular phylogenetic studies (e.g., Tarrio et al., 2001; Da Lage et al., 2007). Using live or alcohol-preserved specimens obtained from stocks of the National Drosophila Species Resource Center at Bowling Green State University (BGSU), the Tucson Fly Stock Center (TFSC) and the Tokyo Metropolitan University (TMU) and the field, we newly determined DNA sequences of *Adh* for seven species, of *28S* for 29 species, of *COII* for 17 species and of *ND2* for 26 species. GenBank accession numbers of these sequences, along with those for known sequences used in the analyses, and the origins of material specimens are given in Table 1.

2.2. DNA extraction, PCR, cloning and sequencing

Total DNA was extracted from a single fly by the standard phenol–chloroform method, or the modified Boom et al.'s (1990) method (Kobayashi et al., 2009). The PCR cycle program comprised an initial 3 min of predenaturation at 94 °C, 35 cycles of amplification (50 s of denaturation at 94 °C; 1 min of annealing at 55 °C for *ND2* and *COII*, 53 °C for *Adh*, and 60 °C for *28S*; 1 min of extension at 72 °C) and 5 min of sequence postextension at 72 °C. The primers (all given left to right from 5′ to 3′ ends) for the PCR and sequencing of each locus are: ND2-FW ATATTTACAGCTTTGAAGG, and ND2-RV AAGCTACTGGGTTCATACC for *ND2* (Wang et al., 2006); COII-FW ATGGCAGATTAGTGCAATGG and COII-RV GTTTAAGAGACCAGTACTTG for *COII* (O'Grady, 1999); 28S-FW CCCGAAGTAT CCTGAATC TTTCGCATTG (Kopp and True, 2002) and 28S-RV TCTTAGTAGCGGCGAGCG (designed by T. Katoh) for *28S*. PCR products were purified with the modified method of Boom et al. (1990).

The *Adh* fragments of *Lordiphosa neokurokawai*, *Lordiphosa vittata*, *Lordiphosa ramula*, *Lordiphosa penicilla*, *Drosophila neocordata* and *Drosophila prosaltans* are amplified by PCR using the primers (from 5′ to 3′ ends) adh-3233 AATGTCTCTCACCAACAAGAAC and adh-3235 AGATGCGCGAGTCTCAGTGCTT (both designed in the present study), and the products were separated by agarose gel electrophoresis. In most cases, multiple bands were presented in the gel, indicating unspecific amplifications. According to GenBank *Adh* sequences of some *Lordiphosa* species, the target region of this gene is highly variable in length among *Lordiphosa* species (ca. 800–1600 bp, primarily due to the length variation of the introns involved). Therefore, multiple bands of desirable length were checked by sequencing the corresponding fragments: each candidate fragment was cloned into the PMD18-T Vector (TaKaRa), then transformed into *Escherichia coli* as host. The recombinant DNA was extracted and sequenced using the ABI 3700 sequencer

Table 1
Samples used for DNA sequencing in the present study, with GenBank accession numbers of sequences (HQ: newly sequenced).

Genus	Subgenus	Species group	Species subgroup	Species	Geographic origin and/or stock[a]	Accession number of sequence			
						Adh	*28S*	*COII*	*ND2*
Drosophila	*Sophophora*	*melanogaster*	*melanogaster*	*melanogaster*	TMU (missing number)	M17833	HQ110532	J01404	U37541
				simulans	TMU (missing number)	M36581	HQ110541	AF474082	AF200854
				yakuba	TMU (missing, from Kenya, Africa)	X54120	HQ110545	X03240	X03240
		obscura	*pseudoobscura*	*pseudoobscura*	BGSU (missing number)	M60979	HQ110539	M95150	HQ110584
				miranda	TFSC (14011–0101.08)	M60998	HQ110533	M95148	HQ110578
		saltans	*elliptica*	*emarginata*	TFSC (14042–0841.04)	AB026526	HQ110528	AF045094	HQ110574
			saltans	*prosaltans*	TFSC (14045–0901.03)	HQ110515,-16,-17	HQ110538	HQ110561	HQ110583
				saltans	TMU (14045–0911.0)	AB026533	HQ110540	HQ110558	HQ110585
			sturtevanti	*sturtevanti*	TFSC (14043–0871.06)	AB026535	HQ110542	HQ110562	HQ110595
			cordata	*neocordata*	TFSC (14041–0831.00)	HQ110512,-13,-14	HQ110535	AF045088	HQ110580
		willistoni	group A	*paulistorum*	TMU (1403–0771.11)	AB026529	HQ110536	HQ110557	HQ110581
				willistoni	TMU (14030–0811.0)	L08648	HQ110544	HQ110560	HQ110587
			group B	*nebulosa*	TMU (14030–0761.05)	HQ110509,-10,-11	HQ110534	AF474099	HQ110579
	Drosophila	*repleta*		*hydei*	BGSU (missing number)	X58694	HQ110530	AF478429	HQ110576
		immigrans	*immigrans*	*immigrans*	BGSU (missing number)	M97638	HQ110531	AF478424	HQ110577
		funebris		*funebris*	TMU (missing number)	AB033643	HQ110529	AF478422	HQ110575
		polychaeta		*polychaeta*	BGSU (missing number)	AB033641	HQ110537	AF478427	HQ110582
		virilis		*virilis*	TMU (missing number)	AB033640	HQ110543	HQ110559	HQ110586
Lordiphosa		*fenestrarum*		*collinella*	Sapporo, Japan	AB026525	HQ110547	HQ110564	HQ110589
		miki		*clarofinis*	Sapporo, Japan	AB026524	HQ110546	HQ110563	HQ110588
				stackelbergi	Sapporo, Japan	AB026534	HQ110552	HQ110569	HQ110594
		nigricolor		*penicilla*	Yunnan, China	HQ110506,-07,-08	HQ110550	HQ110567	HQ110592
		denticeps		*neokurokawai*	Yunnan, China	HQ110518,-19,-20	HQ110549	HQ110566	HQ110591
				vittata	Yunnan, China	HQ110524,-25,-26	HQ110553	HQ110570	HQ110596
				kurokawai	Nagano, Japan	AB026541	HQ110548	HQ110565	HQ110590
				ramula	Yunnan, China	HQ110521,-22,-23	HQ110551	HQ110568	HQ110593
Scaptomyza				*pallida*	Sapporo, Japan	AB033645	HQ110554	HQ110571	HQ110597
Scaptodrosophila		*rufifrons*		*lebanonensis*	TFSC (11010–0021, Lebanon 1733)	X54814	HQ110555	HQ110572	HQ110598
Chymomyza				*procnemis*	TMU (missing number)	AB026521	HQ110527	HQ110556	HQ110573

[a] TMU: stocks maintained at Tokyo Metropolitan University; BGSU, stocks from the National *Drosophila* Species Resource Center at Bowling Green State University; TFSC, stock from the Tucson Fly Stock Center.

following the protocol by the manufacturer. The M13 universal primers AAGCTTGCATGCCTGCAGGTCGACGAT and CGGTACCCGGG GATCCTCTAGAGAT were used for sequencing reaction. Two additional, sequencing primers, adh-3236 CAGTGACGGGATTCAATGCC and adh-3237 ACATCCAGCCAGGAGTTGAA (binding on the exon 2 and 3, respectively; both designed in the present study), were used to determine the long intron II sequence in some species. For each of the seven species, sequences of three different clones were determined, and the consensus of the coding parts of the three sequences was used for phylogenetic analyses.

2.3. Sequence aligning

Newly collected sequences were edited using SeqMan (DNAStar Inc. 1996). The homologous GenBank sequences were downloaded. The concatenated sequences of the four genes, *ND2*, *COII*, *Adh* (only the coding regions) and *28S*, of the 29 species were aligned in MEGA4.1 (Tamura et al., 2007) by the Clustal W method (Thompson et al., 1994). The alignments were then adjusted by eye to make it conform to the codon assignments, with the ends slightly trimmed to reduce the number of end gaps. In addition, the sequences of the three protein-coding genes were translated to amino acid sequences in MEGA4.1.

2.4. Phylogenetic analyses

So far a number of phylogenetic methods, either distance- or model-based (Maximum Parsimony, Maximum Likelihood (ML), etc.), have been developed to account for the among-taxa compositional heterogeneity (Rosenberg and Kumar, 2003, and references therein). However, some of them are impractical for our data. For example, ML algorithms such as Galtier and Gouy's (1998) are computationally too time-consuming in cases including more than seven or eight sequences (Tarrio et al., 2001). Conservative approaches, such as excluding third codon positions (e.g., Delsuc et al., 2002) and/or data partitions that fail in tests for compositional homogeneity (e.g., Springer et al., 1999) and RY-coding (e.g., Phillips et al., 2004), have proved to be effective for increasing the ratio of phylogenetic signal to noise in large datasets such as genome-scale ones. However, these conservative methods will meanwhile result in loss of some phylogenetic signals in cases of smaller datasets such as ours.

Here, we employed the following three methods to account for the compositional heterogeneity in tree construction: modeling of compositional heterogeneity in a Bayesian framework by Foster (2004), the modified Tamura-Nei distance estimation method by Tamura and Kumar (2002), and using of translated amino acid sequences for protein-coding genes. Among the distance-based methods, probably the most popular distance correction for coping with the problem of heterogeneous base composition is the LogDet transformation (Lockhart et al., 1994). However, the LogDet distances are paralinear, i.e., they are expected to show linearity with time and are actually not designed to measure the actual number of substitutions (Lockhart et al., 1994). For instance, it is known that the LogDet method will overestimate evolutionary distances if the four bases do not occur with the equal frequency in the nucleotide sequences compared, even when the evolutionary process is homogeneous (Swofford et al., 1996). Tamura and Kumar (2002) modified the Tamura-Nei method (Tamura and Nei, 1993), which measures the actual number of substitutions irrespective of the base frequency bias, when the evolutionary process is homogeneous. They relaxed the last assumption, i.e., the assumption of the substitution pattern homogeneity among lineages, and demonstrated by computer simulations and empirical data analyses that the modified method performed much better than the LogDet method.

For tree construction, if we use simply concatenated sequences of several protein- and RNA-coding genes, it allows only one substitution model for characters that should be under heterogeneous evolutionary processes. Partitioning such concatenated data into more-homogeneous subsets cancels mal-effects of parameter estimation from smaller data-subsets that reduce the signal-to-variance ratio (Phillips and Penny, 2003). Therefore, we took a sequence-partitioning approach by dividing the concatenated sequences into ten data-subsets: nine by codon position of the three protein-coding genes and one RNA-coding gene, designated as *ND2*-1st, *ND2*-2nd, *ND2*-3rd, *COII*-1st, *COII*-2nd, *COII*-3rd, *Adh*-1st, *Adh*-2nd, *Adh*-3rd and *28S*.

Each of the data-subsets was separately subjected to a test of nucleotide substitution saturation using the index of substitution saturation (I_{ss}) of Xia et al. (2003) implemented in the DAMBE software version 5.2.13 (Xia and Xie, 2001), and a composition homogeneity test using Statio (http://homes.bio.psu.edu/people/faculty/Nei/software.htm) that implements the method of Rzhetsky and Nei (1995) with account of possible phylogenetic correlations.

Before Bayesian analyses using the method of Foster (2004), we first selected the best-fit substitution model for each data-subset using Modeltest 3.7 (Posada and Crandall, 1998). Then, the analyses were conducted under the partitioned model of Bayesian framework, in which parameters of the selected substitution model were applied to each partition (data-subset), using the software p4 (http://bmnh.org/~pf/p4.html). Two MCMC runs of 1,000,000 generations were performed, with sampling of every 100 generations. For each run, 10,001 samples were produced, among which 2500 early-phase ones were discarded as burn-in. Then a majority rule consensus tree showing all compatible taxon bipartitions was obtained by summarizing the remaining 15,002 samples from the two runs (7501 from each). To estimate the least numbers of composition vectors necessary for the model composition to fit the data, different configurations of the numbers of composition vectors were accommodated in the model, starting from the configuration of a single vector (meaning the compositional homogeneity) for every partition. The fit of the model composition to each data partition was assessed by posterior predictive simulation using a tree- and model-based test statistic X^2 (see Foster, 2004 for the calculation procedure). The tail-area probability p_t was calculated by comparing the distribution of X^2s simulated from the two MCMC runs (after burning-in) to the X^2 of the original data-subset. If $p_t < 0.05$, it means that the number of composition vectors accommodated is not enough to adequately model the data. While this requirement ($p_t > 0.05$) is not satisfied in any of the ten data-subsets, more compositional vectors should be added, one by one, for those data-subsets, until the fit of model is achieved for all the data-subsets. Finally, accommodating the least, necessary condition thus determined for the numbers of composition vectors in the model, the Bayesian tree was constructed, and compared with the initial tree under the assumption of compositional homogeneity for all the data-subsets.

For the Neighbor Joining (NJ) tree construction, we first estimated evolutionary distances, without consideration of the among-site variation in the substitution rate (the gamma shape parameter), separately for each data-subset, between the 29 studied species by the MCL (Maximum Composite Likelihood) method of Tamura et al. (2004a, 2007). Then, the resulting ten distance matrices were merged into a single matrix of average evolutionary distances (the number of nucleotide substitutions per site), taking into account the length (the total number of nucleotide sites) of each data-subset. A NJ tree was constructed on the basis of the merged distance matrix. For data-subsets where the compositional homogeneity was rejected by the Statio test, the evolutionary distances were corrected by the method of Tamura and Kumar (2002), allowing the nucleotide composition to vary among the taxa. The resulting tree was compared with that inferred from the assumption of compositional homogeneity in every data-subset. Branch support values were calculated by a bootstrap analysis with 1000 replicates: in each replicate, random sampling of nucleotide sites was made within each data-subset, and a tree was constructed in the above procedure, based on the ten data-subsets obtained by such bootstrap sampling. All analyses were performed using the program MultiGeneNJ developed by Tamura (available from him).

Before constructing a ML tree based on the amino acid sequences, we selected the best-fit substitution model for the concatenated sequences of the three protein fragments, Adh, ND2 and COII, using ProtTest 2.1 (Abascal et al., 2005). The best-fit model selected under the AIC statistical framework was JTT + I + G + F, i.e., the common matrix JTT (Jones et al., 1992) modified with the parameters I (considering that a fraction of the amino acids are invariable), G (including a gamma distribution to account for different rates of change at different positions) and F (using the observed amino acid frequencies). A ML tree was searched in PhyML (Guindon and Gascuel, 2003) under the selected model, with a BioNJ tree calculated also by PhyML as the starting tree, and the proportion of invariable sites (I) and the gamma shape parameter optimized by PhyML. The branch support values were calculated by a non-parametric bootstrap analysis with 500 replicates.

2.5. Estimation of divergence times

Based on the Bayesian tree inferred by modeling the compositional heterogeneity in p4, we estimated the divergence time at each node using the program r8s version 1.71 (Sanderson, 2003). Since no reliable evidence was available for estimating the divergence time between any taxa covered in our analysis, we applied the divergence time of 62.9 Mya between the subgenera *Drosophila* and *Sophophora*, that was estimated by Tamura et al. (2004b) based on a well-established estimate of 5.1 Mya between *Drosophila picticornis* Grimshaw, 1901 endemic to Kauai and its relatives on other islands of Hawaii (Carson and Clauge, 1995), as a calibration point. The penalized likelihood (PL) method was used for divergence time estimation, with a truncated Newton (TN) algorithm for finding optima of the various objective functions. We created 100 bootstrap replicates of our original sequence data with the SeqBoot module in the PHLIP package (Felsenstein, 2004). These bootstrap replicates were run in PAUP*4.0b10 (Swofford, 2003) under the constraint of the selected tree topology, resulting in 100 trees of different branch lengths under the same topology. With the aid of the r8s bootstrap kit (Eriksson, 2007), we checked the cross-validation over a range of discrete values of the smoothing parameter ($10^{-3}, 10^{-2.5}, \ldots, 10^{3.5}$) for the 100 trees in batches in r8s (Sanderson, 2003), and then reran these trees with the level of smoothing selected to estimate the divergence times. One of the outgroup taxa, *Chymomyza procnemis* (Williston, 1896), was pruned from the trees in the r8s analyses to avoid the basal trichotomy. The r8s log file was then scanned for each node across the 100 bootstrap replicates, and the mean of thus estimated 100 divergence times was calculated, along with the standard deviation and the 95% confidence interval after the normal distribution test for the bootstrap estimates.

3. Results

The alignment of the DNA sequences of the four gene regions spanned a total of 2977 nucleotide positions, with the detailed information shown in Table 2. The *28S* region of the alignment was relatively conservative, involving only 23 indels, mostly of 1–3 bases, scattered across the region.

The results of the substitution saturation tests using DAMBE, the composition homogeneity test using Statio, and model selection using Modeltest are shown in Table 3. In the DAMB tests for the whole concatenated sequences of the four genes, the value of substitution saturation index (I_{ss}) was significantly smaller than the critical value, under the assumption of either a symmetrical ($I_{ss.cSym}$) or a very asymmetrical true tree ($I_{ss.cAsym}$). The same resulted from when considering only 1st or 2nd codon positions of each of *ND2*, *COII* and *Adh* genes, as well as the *28S* data-subset, indicating that these data-subsets are unlikely to have experienced saturation. For the three data-subsets consisting of only the 3rd codon positions of each of the protein-coding genes, little saturation was detected as well under the assumption of a symmetrical topology of the true tree. Under the assumption of a very asymmetrical tree topology, however, the I_{ss} values were significantly larger than the critical values ($I_{ss.cAsym}$), indicating poor signals for phylogenetics in these data-subsets, although the assumption is generally very unlikely (Xia and Lemey, 2009). In the compositional homogeneity test, the null hypothesis was rejected, i.e., the nucleotide compositions to be regarded as heterogeneous among the studied taxa, for the whole data, and *ND2*-1st, *ND2*-3rd, *COII*-3rd, *Adh*-1st and *Adh*-3rd data-subsets. The model selected for each data-subset and its parameters shown in Table 3 were incorporated into the partitioned Bayesian analyses.

With respect to the compositional heterogeneity in each data-subset, however, the Bayesian analyses using posterior predictive simulation by MCMC runs showed a slightly different result (Table 4). Of the above five data-subsets, the nucleotide compositions of *ND2*-1st could adequately be modeled by a single composition vector, i.e., an assumption of the compositional homogeneity. On the other hand, in Test 1 where a single vector was applied to every data-subset, the model did not fit to the data, i.e., the compositional homogeneity to be rejected, in the other four data-subsets, as well as in the test by Statio. By increasing the number of vectors, one by one, for these data-subsets in Tests 2 and 3 (Table 4), finally, the least, necessary numbers of vectors were estimated as 2 for *COII*-3rd and *Adh*-1st, and 3 for *ND2*-3rd and *Adh*-3rd.

The Bayesian tree inferred from the partitioned model accommodating this configuration of composition vector numbers is shown in Fig. 1. Its topology was almost the same as those of the other trees (not shown) constructed by the different methods, i.e., the Bayesian tree under the assumption of compositional homogeneity, the NJ trees constructed by the MultiGeneNJ under the assumptions of compositional heterogeneity and homogeneity, and the ML tree based on amino acid sequences of the three protein-coding genes. Most of the internal branches were commonly seen in all the five trees and strongly supported at least in three of them, except for a few weakly supported branches with respect to the relationships of the *nigricolor* group within the genus *Lordiphosa* and among species of the *virilis-repleta* lineage including the genus *Scaptomyza* Hardy (Fig. 1, Table 5).

Table 2
A summary of the alignment of the sequences.

	Nucleotide sequences			Translated amino acid sequences		
	Sites[a]	V[b]	PI[c]	Sites	V	PI
ND2	934	492	374	310	149	115
COII	667	247	181	222	47	15
Adh	711	393	309	237	125	81
28S	665	141	80	–	–	–
Total	2977	1273	944	769	321	211

[a] Number of sites.
[b] Number of variable sites.
[c] Number of parsimony informative sites.

The ingroup was divided into two strongly supported clades at the base of the trees; one consisted of the subgenus *Sophophora* and the genus *Lordiphosa*, but the other of the subgenus *Drosophila* and the genus *Scaptomyza*. Within the former clade, the monophyly of each species group and *Lordiphosa* itself was corroborated, except for the *Lordiphosa nigricolor* and *fenestrarum* groups each of which was represented by a single species. However, *Sophophora* was regarded as paraphyletic in respect to *Lordiphosa*. The latter was placed as the sister to the Neotropical *Sophophora* consisting of the *saltans* and *willistoni* groups, and the clade of *Lordiphosa* + Neotropical *Sophophora* was sister to that of the *obscura* and *melanogaster* groups. Within *Lordiphosa*, relationships between the species groups were uncertain, except for the sister relationship between the *fenestrarum* and *miki* groups. Taxon sampling was too scarce to note any phylogenetic relationships within the clade of *Drosophila s. str.* and *Scaptomyza* in the present study.

Divergence times were estimated for the nodes of the Bayesian tree shown in Fig. 1, and the results are represented in Table 5. The earliest split, by the genus *Scaptodrosophila*, at the root of the tree was estimated to have occurred very old, 74.1 ± 2.9 (mean ± s.d.) Mya, in the Cretaceous Period. Within the sophophoran clade including *Lordiphosa*, divergence between the *obscura* + *melanogaster* and the *Lordiphosa* + Neotropcical *Sophophora* clades (at node C) was estimated to have taken place 57.3 ± 1.7 Mya in the Paleocene Epoch. The event in which we are most interested is the divergence between *Lordiphosa* and Neotropcical *Sophophora* (at node H), and it was estimated that this disjunction of the Old World and the New World tropical lineages occurred 45.9 ± 1.9 Mya in the middle Eocene Epoch. Diversification of species groups within *Sophophora* and *Lordiphosa* was estimated to have not synchronously occurred in different lineages; the *obscura* and the *melanogaster* groups diverged from each other (at node G) in the middle Eocene (46.6 ± 2.3 Mya), the *saltans* and the *willistoni* groups (at node K) in the middle Oligocene (29.1 ± 2.0 Mya), and the species groups of *Lordiphosa* diversified (at nodes L, O and R) in the late Oligocene (25.5 Mya) to the middle Miocene (16.3 Mya).

4. Discussion

In this study, we analyzed the phylogeny of the subgenus *Sophophora*, especially focusing on its relationship to the genus *Lordiphosa* (Katoh et al., 2000; Hu and Toda, 2001). In the analyses, we paid much attention to the problem of among-taxa nucleotide compositional heterogeneity, which has proved to compromise the phylogeny reconstruction in *Sophophora* with a large variation in GC content among different species groups (Tarrio et al., 2001; Da Lage et al., 2007). In addition, we took a sequence-partitioning approach to allow multiple substitution models for nucleotide sequences that have evolved under different evolutionary processes. In particular, we developed a novel, sequence-partitioning procedure (the program MultiGeneNJ) for NJ tree construction and bootstrap analysis. In consequence, however, the trees inferred from the different methods and assumptions were almost identical in topology and branch supports. This implies that phylogenetic signals involved in our data-set are so strong as not to be overwhelmed by the compositional heterogeneity, which is really present among the nucleotide sequences of our data, as well as in many other cases (e.g., Conant and Lewis, 2001; Rosenberg and Kumar, 2003). The efficiency of our novel method for NJ tree construction should be tested in more delicate cases and/or simulated phylogenies.

The most remarkable inference from this study is that *Sophophora* is paraphyletic: *Lordiphosa* is placed as the sister to the Neotropical *Sophophora* consisting of the *saltans* and *willistoni* groups, and *Sophophora* is divided into the clade of *Lordiphosa* +

Table 3
Results of substitution saturation tests (using DAMBE), composition homogeneity tests (using Statio), and model selection (using Modeltest 3.7).

	Whole concatenated	*ND2*-1st	*ND2*-2nd	*ND2*-3rd	*COII*-1st	*COII*-2nd	*COII*-3rd	*Adh*-1st	*Adh*-2nd	*Adh*-3rd	*28S*
Test of substitution saturation											
I_{ss}	0.1772	0.2041	0.0904	0.5151	0.0754	0.0209	0.4581	0.1690	0.0788	0.5532	0.0609
$I_{ss.cSym}$[a]	0.6602	0.6851	0.6852	0.6839	0.6742	0.6740	0.6740	0.6758	0.6780	0.6759	0.7235
P_{Sym}[b]	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$I_{ss.cAsym}$[c]	0.3421	0.3782	0.3784	0.3764	0.3623	0.3621	0.3621	0.3646	0.3747	0.3648	0.4409
P_{Asym}[d]	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0015	0.0000	0.0000	0.0000	0.0000
Compositional homogeneity test											
df	84	84	84	84	84	84	84	84	84	84	84
I	300.24	146.25	96.33	266.21	62.91	13.88	286.30	151.54	65.79	487.84	73.76
P	0.0000[e]	0.0000[e]	0.1687	0.0000[e]	0.9587	1.0000	0.0000[e]	0.0000[e]	0.9291	0.0000[e]	0.7801
Substitution model selection											
Model selected	GTR + I + G	GTR + I + G	HKY + I + G	HKY + I + G	TrNef + I + G	F81 + I + G	HKY + I + G	TrN + G	F81 + G	TrN + G	TVM + I + G
Nucleotide frequency											
A	0.3254	0.4030	0.2382	0.4907	Equal	0.2684	0.4355	0.3572	0.3206	0.1883	0.3598
C	0.1532	0.0494	0.1950	0.0468	Equal	0.1923	0.0533	0.2136	0.2746	0.3148	0.1181
G	0.1505	0.1159	0.1173	0.0288	Equal	0.1365	0.0154	0.2976	0.1524	0.2220	0.1709
T	0.3709	0.4318	0.4495	0.4338	Equal	0.4029	0.4958	0.1316	0.2523	0.2748	0.3511
Ti/tv ratio	–	–	1.2227	5.5401	–	–	34.5426	–	–	–	–
Rate matrix											
A–C	3.1000	13.7540	–	–	1.0000	Equal	–	1.0000	Equal	1.0000	2.4791
A–G	7.5824	8.1949	–	–	7.5647	Equal	–	1.3148	Equal	2.7809	6.8025
A–T	5.0865	3.3064	–	–	1.0000	Equal	–	1.0000	Equal	1.0000	2.8241
C–G	5.0004	10.0123	–	–	1.0000	Equal	–	1.0000	Equal	1.0000	0.0638
C–T	14.2760	96.7363	–	–	19.2390	Equal	–	6.5109	Equal	4.0536	6.8025
G–T	1.0000	1.0000	–	–	1.0000	Equal	–	1.0000	Equal	1.0000	1.0000
I[f]	0.4616	0.3693	0.5332	0.0198	0.6559	0.8559	0.0660	0	0	0	0.6100
α[g]	1.0478	0.7826	0.6162	0.7057	1.0152	0.7422	0.4078	0. 4447	0.2499	2.5882	0.5934

[a] Index of substitution saturation assuming a symmetrical true tree.
[b] Probability of significant difference between I_{ss} and $I_{ss.cSym}$ (two-tailed test).
[c] Index of substitution saturation assuming an asymmetrical true tree.
[d] Probability of significant difference between I_{ss} and $I_{ss.cAsym}$ (two-tailed test).
[e] Rejecting the null hypothesis, i.e., the compositional homogeneity.
[f] Proportion of invariable sites.
[g] Gamma distribution shape parameter.

Table 4
Results of Bayesian composition fit tests by posterior predictive simulations. The tail-area probability (p_t) was calculated for each data partition using the post-burn-in values of the composition X^2 statistic from simulations in two replicated MCMC runs.

Test	Numbers of composition vectors (given in turn for partitions 1–10)[a]	p_t for each partition									
		1 (*ND2*-1st)	2 (*ND2*-2nd)	3 (*ND2*-3rd)	4 (*COII*-1st)	5 (*COII*-2nd)	6 (*COII*-3rd)	7 (*Adh*-1st)	8 (*Adh*-2nd)	9 (*Adh*-3rd)	10 (*28S*)
1	(1, 1, **1**, 1, 1, **1**, **1**, 1, **1**, 1)	0.1864	0.3790	0.0000*	0.8453	0.8831	0.0003*	0.0000*	0.3462	0.0000*	0.3337
2	(1, 1, **2**, 1, 1, 2, 2, 1, **2**, 1)	0.2723	0.4181	0.0279*	0.8814	0.9004	0.0604	0.1082	0.3305	0.0244*	0.3608
3	(1, 1, 3, 1, 1, 2, 2, 1, 3, 1)	0.1760	0.4416	0.0823	0.8860	0.8643	0.0804	0.1425	0.3177	0.0615	0.3834

[a] The bold number.
* Indicate that the number of composition vectors is not enough to adequately model the data.

Neotropical *Sophophora* and the clade of the *obscura* + *melanogaster* groups. This phylogenetic inference for the sophophoran clade including *Lordiphosa* has provisionally been confirmed by a cladistic analysis with morphological characters as well (Hu and Toda, unpublished). Based on the results of such an analysis, the taxonomy of this clade will be revised elsewhere by designating synapomorphies as the diagnosis for each of the subclades within it.

Our estimation suggested that the outgroup *Scaptodrosophila* diverged from the main body of Drosophilinae in a very old time, approximately 74 Mya, in the Cretaceous Period. This estimate is near but within the maximum possible age of the Drosophilidae, 80 Mya, inferred from geologic and biogeographic evidence: New Zealand that separated from Australia approximately 80 Mya has no endemic drosophilid fauna, while Australia has a well-developed endemic one (Beverley and Wilson, 1984). To estimate the time of origin of Drosophilidae, more evidence is needed, for example from a reliable phylogeny including the subfamily Steganinae, or more desirably from fossils of the Cretaceous Period; the known oldest drosophilid fossil is of *Eletrophortica* Hennig found from upper (late) Eocene strata of 40 Mya (Hennig, 1960).

Throckmorton (1975) depicted two major patterns in the evolution of Drosophilinae. One is the primary disjunction of tropical fauna and the other is the secondary one of temperate fauna between the Old and New Worlds. Both patterns are seen in parallel along several lineages. The vicariant divergence between *Lordiphosa* and the Neotropical *Sophophora* is one of cases showing the first pattern; the *saltans* and *willisoni* groups are endemic to the Neotropics, while *Lordiphosa* is distributed in the Oriental and Palearctic Regions, with the highest species richness in the Oriental Region (De and Gupta, 1996; Gupta and De, 1996; Gupta and Gupta, 1991; Kumar and Gupta, 1990; Okada, 1966, 1984, 1988; Zhang, 1993a,b; Zhang and Liang, 1992, 1994; Toda, unpublished). Throckmorton (1975) suggested that parallel patterns are seen as well in "the *Scaptodrosophila* radiation", "the *virilis-repleta* radiation", "the *immigrans* radiation" and "the *Hirtodrosophila* radiation" of his sense. However, the strict sister relationship between the Old

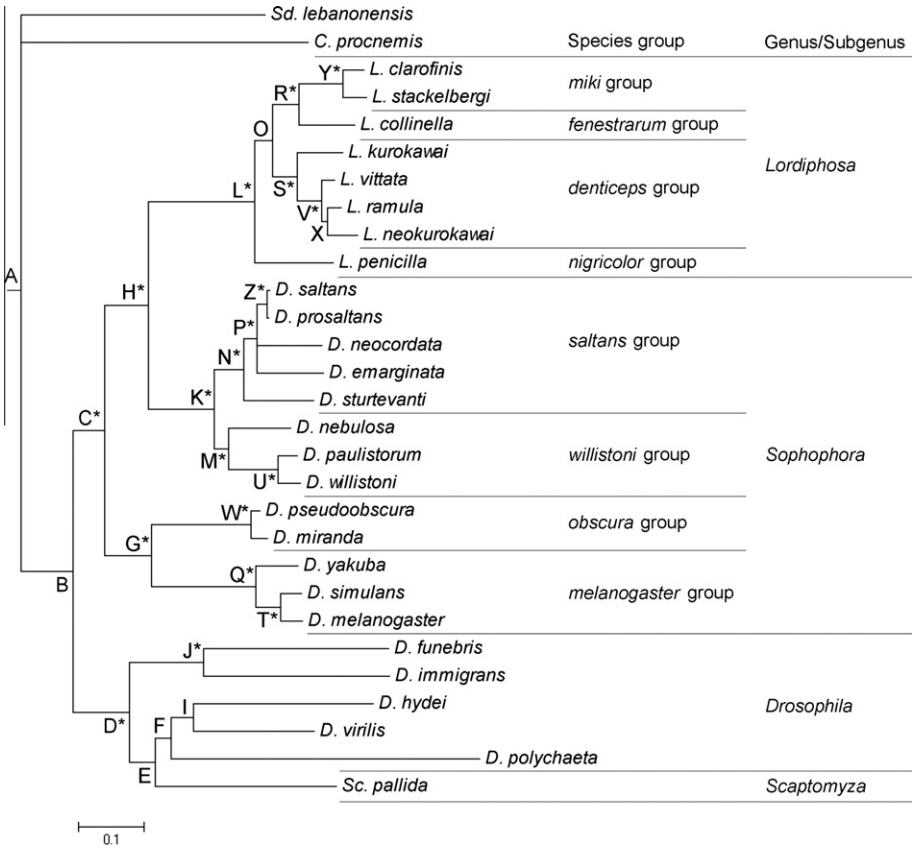

Fig. 1. Bayesian tree constructed under a partitioned model accommodating the configuration of nucleotide composition vector numbers inferred from the posterior predictive simulation (Test 3 in Table 4), using the software p4 (Foster, 2004). * Strongly supported branch (see Table 5).

Table 5
Branch support values (the posterior probability for Bayesian trees and the bootstrap% for NJ and ML trees) and divergence times estimated for the nodes of the Bayesian tree inferred from the model accommodating the compositional heterogeneity. Branch/node codes correspond to those shown in Fig. 1; bold codes indicate internal branches commonly seen in all the five trees and strongly supported (bold numerals) at least in three of them.

Branch/Node	Clade	Bayesian tree		NJ tree		ML tree	Divergence time (Mya)		
		C_{hetero}[a]	C_{homo}[a]	C_{hetero}[a]	C_{homo}[a]	AA[a]	Mean	95% confidence interval	Geological age
A	(Root)	–	–	–	–	–	74.1	73.6–74.7	Cretaceous
B	(Ingroup)	–	–	–	–	–	62.9[b]	–	
C	*Sophophora* + *Lordiphosa*	**1.00**	**1.00**	**81**	**81**	**97**	57.3	57.0–57.7	Paleocene
D	*Drosophila* + *Scaptomyza*	**1.00**	**1.00**	**96**	**96**	**98**	54.4	54.0–54.7	Eocene
E		1.00	1.00	–	–	90	49.5	49.1–49.9	
F		0.73	0.92	–	39	80	48.2	47.7–48.7	
G	*obscura* gr.+*melanogaster* gr.	**1.00**	**1.00**	**99**	**99**	**99**	46.6	46.1–47.0	
H	*Lordiphosa* + Neotropcial *Sophophora*	**1.00**	**1.00**	**99**	**99**	65	45.9	45.5–46.2	
I		0.50	–	–	–	–	41.9	41.4–42.3	
J		**1.00**	**1.00**	**90**	**91**	**98**	40.8	40.3–41.3	
K	*saltans* gr.+*willistoni* gr.	**1.00**	**1.00**	**100**	**100**	**99**	29.1	28.7–29.4	Oligocene
L	*Lordiphosa*	**1.00**	**1.00**	**100**	**100**	**100**	25.5	25.2–25.9	
M	*willistoni* gr.	**1.00**	**1.00**	**92**	**92**	50	24.1	23.7–24.5	
N	*saltans* gr.	**1.00**	**1.00**	**97**	**98**	**87**	22.3	21.9–22.7	Miocene
O		0.98	0.84	–	61	–	22.1	21.7–22.4	
P		**1.00**	**1.00**	57	55	**91**	19.9	19.5–20.3	
Q	*melanogaster* gr.	**1.00**	**1.00**	**100**	**100**	**100**	18.9	18.5–19.3	
R	*miki* gr.+*fenestrarum* gr.	**1.00**	**1.00**	**98**	**98**	**90**	16.3	15.9–16.6	
S	*denticeps* gr.	**1.00**	**1.00**	**100**	**100**	**100**	16.2	15.9–16.5	
T		**1.00**	**1.00**	**100**	**100**	52	9.9	9.6–10.3	
U		**1.00**	**1.00**	**100**	**100**	**98**	8.8	8.6–9.0	
V		**1.00**	**1.00**	**100**	**100**	**100**	8.4	8.2–8.6	
W	*obscura* gr.	**1.00**	**1.00**	**100**	**100**	**100**	7.5	7.3–7.8	
X		0.93	0.94	–	–	88	7.3	7.1–7.5	
Y	*miki* gr.	**1.00**	**1.00**	**100**	**100**	**100**	6.3	6.1–6.5	
Z		**1.00**	**1.00**	**100**	**100**	**99**	1.8	1.7–2.0	Pleistocene

[a] The tree was constructed under the assumption of compositional heterogeneity (C_{hetero}) or compositional homogeneity (C_{homo}), or based on the amino acid sequences (AA).
[b] The calibration point adopted from the estimation by Tamura et al. (2004b).

and New World clades has not been corroborated yet in any of these cases. In addition, Throckmorton (1975) considered that the tropical disjunction of these lineages had occurred in the early Oligocene times. However, estimation by molecular methodologies based on protein immunological distances (Beverley and Wilson, 1984) and genomic mutation distances (Tamura et al., 2004b) made the tropical disjunction of the sophophoran lineage trace back to the early Eocene (53 Mya) and even to the Paleocene (62.2 Mya), respectively, although both studies estimated the divergence time between *D. melanogaster* and *D. willisoni*. The present study has revealed that the sister group to the Neotropical *Sophophora* is *Lordiphosa*. And, based on this finding, the tropical disjunction, between *Lordiphosa* and the Neotropical *Sophophora*, in the sophophoran lineage was estimated to have occurred in the younger age, the middle Eocene (45.9 Mya), than the two estimates by other molecular methods, but still older than the age postulated by Throckmorton (1975). Furthermore, this estimate falls very close to the time (46 Mya) of *virilis-repleta* divergence, another case of the tropical disjunction between the Old and New Worlds, estimated by Beverley and Wilson (1984). Reasonably, our estimate (57.3 Mya) for the time of divergence between the *obscura* + *melanogaster* and the *Lordiphosa* + Neotropcical *Sophophora* clades is not so incongruent with the two previous estimates for the *melanogaster–willisoni* divergence, being just intermediate between them.

Based on the estimated time of divergence between *Lordiphosa* and the Neotropical *Sophophora* and evidence from paleontology, paleo-geography and -climatology, a scenario of the disjunction of tropical fauna between the Old and New Worlds in an early phase of drosophilid evolution is drawn as follows. The period (from 57.3 to 45.9 Mya) during which the common ancestor of *Lordiphosa* and the Neotropical *Sophophora* is estimated to have existed corresponds largely to the Paleocene–Eocene (or Late Paleocene) Thermal Maximum and the Early Eocene Climatic Optimum with highest temperatures of the Earth's surface in the Cenozoic (Zachos et al., 2001). In this period, the Earth was entirely covered with forests from pole to pole, apart from the driest deserts. A tropical and/or paratropical biota was continuously distributed throughout the continents, from Asia through Europe to North America, of the Northern Hemisphere (Scotese, 2002), with thermophilic plants and animals distributed even to high latitudes, 50–60°N (Tiffney and Manchester, 2001). Plentiful fossil records of plants, mammals and some insects from the Paleocene and Eocene strata well document the cross-North Atlantic biogeographic pattern, strongly suggesting dispersal of the tropical biota between Europe and North America via Greenland (Archibald et al., 2006, and references therein; Eldholm and Thiede, 1980; McKenna, 1975). The presence of all major skeleton lineages of the Drosophilidae in the Old World (Throckmorton, 1975), at present well conserved in the Oriental Region (Okada, 1981), suggests its origin and early radiations in the Old World tropics. If it is the case, the common ancestor of *Lordiphosa* and the Neotropical *Sophophora* as well should have originated in the tropical forest nearly continuously stretching from southern Asia to Europe, in the late Paleocene according to our divergence time estimation. However, there is geologic evidence that the Asian and European faunas had been more or less separated from each other by the Turgai Strait having persisted in the present-day West Siberia and functioned as a dispersal barrier until the Oligocene (Akhmetiev and Beniamovski, 2006; Tiffney and Manchester, 2001, and references therein). On the other hand, some plant fossil records from the north shore of the Tethys, though fragmentary, indicate the continuity of tropical or subtropical vegetation from southern Asia (China) to Europe through a narrow corridor along the shore (Tiffney and Manchester, 2001). Thus, the circumstantial evidence suggests that the common ancestor of *Lordiphosa* and the Neotropical *Sophophora* dispersed from Europe to North America via the North Atlantic Land Bridge in the late Paleocene or the early Eocene. Although its dispersal from Asia to North America via the Bering Land Bridge can not absolutely be ruled out, some barriers may have made it difficult for tropical biota to cross this route. The tropical biota of southern Asia had been separated by an arid zone in central China during the Paleocene to Eocene period (Tiffney and Manchester, 2001), from northeastern Asia with a temperate biota which was under the influence of a cold current running along Northeastern Pacific coasts (Akhmetiev, 2007). Another possible adverse condition for the dispersal of tropical biota may have resulted from the more northerly location of the Bering Land Bridge at that time than at the present; it lay at approximately 80°N, which placed the bridge in an area with extended darkness and cold in winter and would have made it difficult for tropical biota such as evergreen angiosperms and animals without diapausing ability to cross the bridge (Tiffney and Manchester, 2001). The estimated time, 45.9 Mya, of the divergence between *Lordiphosa* and the Neotropical *Sophophora* is near and just after the time, 49 Mya, of the North Atlantic Land Bridge break estimated from mammalian fossils (McKenna, 1975), thus supporting the hypothesis of the dispersal from Europe to North America in the common ancestor of *Lordiphosa* and the Neotropical *Sophophora*. The disjunction of tropical biota between Europe and North America in the middle Eocene should have primarily been caused by the tectonic force spreading the sea floor that broke directly the physical connection between Europe and Greenland (Eldholm and Thiede, 1980). Additionally, the global cooling in this period (Zachos et al., 2001) as well should have contributed to this disjunction causing southerly retreat of tropical forest and its fragmentation between the Old and New Worlds. The latter process coincides with the pattern found by Tamura et al. (2004b) that some clusters of divergence events in the evolution of Drosophilidae correspond to the periods of major Cenozoic climatic cooling. The proposed hypothesis, i.e., the dispersal of tropical drosophilid fauna from Europe to North America in the late Paleocene to the early Eocene and its disjunction between them in the middle Eocene, should be tested, based on reliable reconstruction of phylogenies and estimation of divergence times, for genuine sister groups of other drosophilid lineages showing parallel vicariant biogeographic patterns between the Old and New World tropics.

Acknowledgments

We thank Dr. P. Foster of the Natural History Museum, London for valuable suggestions on using of the program p4, Dr. T. Chigai of Hokkaido University (HU) for valuable discussion in developing the novel procedure of NJ tree construction, Prof. T. Aotsuka and Mr. T. Ide of Tokyo Metropolitan University for the material support, Dr. K. Shimada of HU for kind technical guidance in DNA sequencing, Dr. Y. Kobayashi of HU for helpful suggestion of relevant literature of the geology and paleontology, and Dr. T. Eriksson of the Royal Swedish Academy of Sciences for helpful suggestions on using of the r8s bootstrap kit. This work was partly supported by Grants-in-Aid for JSPS Fellows (No. 16·04184) and for Scientific Research (No. 20570080) from the Japan Society for the Promotion of Science to Yao-guang Hu, and the National Basic Research Program of China (No. 2007CB411600, as a participator) and NSFC (No. 30760037) to Jian-jun Gao.

References

Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21, 2104–2105.

Akhmetiev, M.A., 2007. Paleocene and Eocene floras of Russia and adjacent regions: climatic conditions of their development. Paleontol. J. 41, 1032–1039.

Akhmetiev, M.A., Beniamovski, V.N., 2006. The Paleocene and Eocene in the Russian part of West Eurasia. Stratigr. Geol. Correl. 14, 49–72.
Anderson, C.L., Carew, E.A., Powell, J.R., 1993. Evolution of the *Adh* locus in the *Drosophila willistoni* group: the loss of an intron and shift in codon usage. Mol. Biol. Evol. 10, 605–618.
Archibald, S.B., Cover, S.P., Moreau, C.S., 2006. Bulldog ants of the Eocene Okanagan Highlands and history of the subfamily (Hymenoptera: Formicidae: Myrmeciinae). Ann. Entomol. Soc. Am. 99, 487–523.
Beverley, S.M., Wilson, A.C., 1982. Molecular evolution in *Drosophila* and higher Diptera. I. Micro-complement fixation studies of a larval hemolymph protein. J. Mol. Evol. 18, 251–264.
Beverley, S.M., Wilson, A.C., 1984. Molecular evolution in *Drosophila* and higher Diptera. II. A time scale for fly evolution. J. Mol. Evol. 21, 1–13.
Boom, R., Sol, C.J.A., Salimans, M.M.M., Jansen, C.L., Wertheim-Van Dillen, P.M.E., Van der Noordaa, J., 1990. Rapid and simple method for purification of nucleic acids. J. Clin. Microbiol. 28, 495–503.
Burla, H., 1954. Zur Kenntnis der Drosophiliden der Elfenbeinküste. Rev. Suisse Zool. 61 (Suppl.), 1–218.
Carson, H.L., Clauge, D.A., 1995. Geology and biogeography of Hawaii. In: Wagner, W.L., Funk, V.A. (Eds.), Hawaiian Biogeography: Evolution on a Hot Spot Archipelago. Smithsonian Institution Press, Washington, D.C, pp. 14–29.
Conant, G.C., Lewis, P.O., 2001. Effects of nucleotide composition bias on the success of the parsimony criterion in phylogenetic inference. Mol. Biol. Evol. 18, 1024–1033.
Da Lage, J.L., Kergoat, G.L., Maczkowlak, F., Silvain, J.F., Cariou, M.L., Lachase, D., 2007. A phylogeny of Drosophilidae using the *Amyrel* gene: qustioning the *Drosophila melanogaster* species group boundaries. J. Zoolog. Syst. Evol. Res. 45, 47–63.
De, A., Gupta, J.P., 1996. Records of drosophilid species from West Bengal with description of one new and two previously unrecorded species from India (Insecta: Diptera: Drosophilidae). Senckenb. Biol. 76, 129–133.
Delsuc, F., Scally, M., Madsen, O., Stanhope, M.J., de Jong, W.W., Catzeflis, F.M., Springer, M.S., Douzery, E.J.P., 2002. Molecular phylogeny of living xenarthrans and the impact of character and taxon sampling on the placental tree rooting. Mol. Biol. Evol. 19, 1656–1671.
Dobzhansky, T., Powell, J.R., 1975a. *Drosophila pseudoobscura* and its American relatives, *Drosophila persimilis* and *Drosophila miranda*. In: King, R.C. (Ed.), Handbook of Genetics: Invertebrates of Genetic Interest, vol. 3. Plenum Press, New York and London, pp. 537–587.
Dobzhansky, T., Powell, J.R., 1975b. The *willistoni* group of sibling species of *Drosophila*. In: King, R.C. (Ed.), Handbook of Genetics: Invertebrates of Genetic Interest, vol. 3. Plenum Press, New York and London, pp. 589–622.
Drosophila 12 Genomes Consortium, 2007. Evolution of genes and genomes on the Drosophila phylogeny. Nature 450, 203–218.
Ehrman, L., Powell, J.R., 1982. The *Drosophila willistoni* species group. In: Ashburner, M., Carson, H.L., Thompson, J.N. (Eds.), The Genetics and Biology of Drosophila, vol. 3b. Academic Press, London, pp. 193–225.
Eldholm, O., Thiede, J., 1980. Cenozoic continental separation between Europe and Greenland. Palaeogeogr. Palaeoclimatol. Palaeoecol. 30, 243–259.
Eriksson, T., 2007. r8s bootstrap kit. Published by the author.
Felsenstein, J., 2004. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.
Foster, P.G., 2004. Modeling compositional heterogeneity. Syst. Biol. 53, 485–495.
Galtier, N., Gouy, M., 1998. Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. Mol. Biol. Evol. 15, 871–879.
Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.
Gupta, J.P., De, A., 1996. Records of Drosophilidae with description of two new species from Bhutan (Insecta: Diptera). Entomon 21, 177–186.
Gupta, K.K., Gupta, J.P., 1991. Four new and two unrecorded species of Drosophilidae from India (Insecta: Diptera). Proc. Zool. Soc., Calcutta 44, 119–126.
Hennig, W., 1960. Die Dipteren-Fauna von Neuseeland als systematisches und tiergeographisches Problem. Beitr. Entomol. 10, 221–329.
Hu, Y.G., Toda, M.J., 2001. Polyphyly of *Lordiphosa* and its relationships in Drosophilinae (Diptera: Drosophilidae). Syst. Entomol. 26, 15–31.
Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. 8, 275–282.
Katoh, T., Tamura, K., Aotsuka, T., 2000. Phylogenetic position of the subgenus *Lordiphosa* of the genus *Drosophila* (Diptera: Drosophilidae) inferred from alcohol dehydrogenase (*Adh*) gene sequences. J. Mol. Evol. 51, 122–130.
Kobayashi, N., Ohta, Y., Katoh, T., Kahono, S., Hartini, S., Katakura, H., 2009. Molecular phylogenetic analysis of three groups of Asian epilachnine ladybird beetles recognized by the female internal reproductive organs and modes of sperm transfer. J. Nat. Hist. 43, 1637–1649.
Kopp, A., True, J.R., 2002. Phylogeny of the oriental *Drosophila melanogaster* species group: a multilocus reconstruction. Syst. Biol. 51, 786–805.
Kumar, A., Gupta, J.P., 1990. Four new species of Drosophilidae (Diptera: Insecta) from Sikkim, India. Proc. Zool. Soc., Calcutta 43, 25–30.
Kwiatowski, J., Ayala, F.J., 1999. Phylogeny of *Drosophila* and related genera: conflict between molecular and anatomical analyses. Mol. Phylogenet. Evol. 13, 319–328.
Kwiatowski, J., Skarecky, D., Bailey, K., Ayala, F.J., 1994. Phylogeny of *Drosophila* and related genera inferred from the nucleotide sequence of the Cu, Zn Sod gene. J. Mol. Evol. 38, 443–454.
Kwiatowski, J., Krawczyk, M., Jaworski, M., Skarecky, D., Ayala, F.J., 1997. Erratic evolution of glycerol-3-phosphate dehydrogenase in *Drosophila*, *Chymomyza* and *Ceratitis*. J. Mol. Evol. 44, 9–22.
Lemeunier, F., Ashburner, M., 1976. Relationships within the *melanogaster* species subgroup of the genus *Drosophila* (*Sophophora*). II. Phylogenetic relationships between six species based upon polytene chromosome banding sequences. Proc. R. Soc. Lond., B, Biol. Sci. 193, 275–294.
Lockhart, P.J., Steel, M.A., Hendy, M.D., Penny, D., 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Mol. Biol. Evol. 11, 605–612.
Mather, W.B., 1955. The genus *Drosophila* (Diptera) in eastern Queensland. I. Taxonomy. Aust. J. Zool. 3, 545–582.
McEvey, S.F., 2009. Taxonomic review of the Australian *Drosophila setifemur* species group, a new name for the *D. dispar* species group (Diptera: Drosophilidae). Rec. Aust. Mus. 61, 31–38.
McKenna, M.C., 1975. Fossil mammals and early Eocene North Atlantic land continuity. Ann. Mo. Bot. Gard. 62, 335–353.
O'Grady, P.M., 1999. Reevaluation of phylogeny in the *Drosophila obscura* species group based on combined analysis of nucleotide sequences. Mol. Phylogenet. Evol. 12, 124–139.
O'Grady, P.M., DeSalle, R., 2008. Out of Hawaii: the origin and biogeography of the genus *Scaptomyza* (Diptera: Drosophilidae). Biol. Lett. 4, 195–199.
O'Grady, P.M., Kidwell, M.G., 2002. Phylogeny of the subgenus *Sophophora* (Diptera: Drosophilidae) based on combined analysis of nuclear and mitochondrial sequences. Mol. Phylogenet. Evol. 22, 442–453.
Okada, T., 1966. Diptera from Nepal, Cryptochaetidae, Diastatidae and Drosophilidae. Bull. Br. Mus., Nat. Hist. 6 (Suppl.), 1–129.
Okada, T., 1981. Oriental species, including New Guinea. In: Ashburner, M., Carson, H.L., Thompson, J.N. (Eds.), The Genetics and Biology of Drosophila, vol. 3a. Academic Press, London, pp. 261–289.
Okada, T., 1984. New or little known species of *Drosophila* (*Lordiphosa*) with taximetrical analyses (Diptera, Drosophilidae). Kontyû 52, 565–575.
Okada, T., 1988. Family Drosophilidae (Diptera) from the Lund University Ceylon expedition in 1962 and Borneo collection in 1978–1979. Entomol. Scand., Suppl. 30, 111–151.
Pélandakis, M., Solignac, M., 1993. Molecular phylogeny of *Drosophila* based on ribosomal RNA sequences. J. Mol. Evol. 37, 525–543.
Pélandakis, M., Higgins, D.G., Solignac, M., 1991. Molecular phylogeny of the subgenus *Sophophora* of *Drosophila* derived from large subunit of ribosomal RNA sequences. Genetica 84, 87–94.
Phillips, M.J., Penny, D., 2003. The root of the mammalian tree inferred from whole mitochondrial genomes. Mol. Phylogenet. Evol. 28, 171–185.
Phillips, M.J., Delsuc, F., Penny, D., 2004. Genome-scale phylogeny and the detection of systematic biases. Mol. Biol. Evol. 21, 1455–1458.
Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817–818.
Powell, J.R., DeSalle, R., 1995. *Drosophila* molecular phylogenies and their uses. Evol. Biol. 28, 87–138.
Remsen, J., O'Grady, P.M., 2002. Phylogeny of Drosophilinae (Diptera: Drosophilidae), with comments on combined analysis and character support. Mol. Phylogenet. Evol. 24, 249–264.
Rodriguez-Trelles, F., Tarrio, R., Ayala, F.J., 1999a. Molecular evolution and phylogeny of the *Drosophila saltans* species group inferred from the *Xdh* gene. Mol. Phylogenet. Evol. 13, 110–121.
Rodriguez-Trelles, F., Tarrio, R., Ayala, F.J., 1999b. Switch in codon bias and increased rates of amino acid substitution in the *Drosophila saltans* species group. Genetics 153, 339–350.
Rosenberg, M.S., Kumar, S., 2003. Heterogeneity of nucleotide frequencies among evolutionary lineages and phylogenetic inference. Mol. Biol. Evol. 20, 610–621.
Rzhetsky, A., Nei, M., 1995. Tests of applicability of several substitution models for DNA sequence data. Mol. Biol. Evol. 12, 131–151.
Sanderson, M.J., 2003. R8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19, 301–302.
Scotese, C.R., 2002. <http://www.scotese.com>, (PALEOMAP website).
Springer, M.S., Amrine, H.M., Burk, A., Stanhope, M.J., 1999. Additional support for Afrotheria and Paenungulata, the performance of mitochondrial versus nuclear genes, and the impact of data partitions with heterogeneous base composition. Syst. Biol. 48, 65–75.
Sturtevant, A.H., 1942. I. The classification of the genus *Drosophila*, with descriptions of nine new species. Univ. Texas Publ. No. 4213, 5–51.
Swofford, D.L., 2003. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4. Sinauer Associates, Sunderland, Massachusetts.
Swofford, D.L., Olsen, G.J., Waddell, P.J., Hillis, D.M., 1996. Phylogenetic inference. In: Hillis, D.M., Moritz, C., Marle, B.K. (Eds.), Molecular Systematics, second edition. Sinauer Associates, Sunderland, Mass., pp. 407–514.
Tamura, K., Kumar, S., 2002. Evolutionary distance estimation under heterogeneous substitution pattern among lineages. Mol. Biol. Evol. 19, 1727–1736.
Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10, 512–526.
Tamura, K., Nei, M., Kumar, S., 2004a. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc. Natl. Acad. Sci. U.S.A. 101, 11030–11035.
Tamura, K., Subramanian, S., Kumar, S., 2004b. Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. Mol. Biol. Evol. 21, 36–44.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596–1599.

Tarrio, R., Rodriguez-Trelles, F., Ayala, F.J., 2001. Shared nucleotide composition biases among species and their impact on phylogenetic reconstructions of the Drosophilidae. Mol. Biol. Evol. 18, 1464–1473.

Tatarenkov, A., Zurovcova, M., Ayala, F.J., Peigler, R.S., 2001. *Ddc* and *amd* sequences resolve phylogenetic relationships of *Drosophila*. Mol. Phylogenet. Evol. 20, 321–325.

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

Throckmorton, L.H., 1975. The phylogeny, ecology and geography of *Drosophila*. In: King, R.C. (Ed.), Handbook of Genetics: Invertebrates of Genetic Interest, vol. 3. Plenum Press, New York and London, pp. 421–469.

Tiffney, B.H., Manchester, S.R., 2001. The use of geological and paleontological evidence in evaluating plant phylogeographic hypotheses in the Northern Hemisphere Tertiary. Int. J. Plant Sci. 162 (6 Suppl.), S3–S17.

Tsacas, L., 1980. Les espèces montagnardes afrotropicales de Drosophilidae (Diptera) I. – Le groupe *Drosophila dentissima*. Ann. Soc. Entomol. Fr. (N.S.) 16, 517–540.

van der Linde, K., Houle, D., 2008. A supertree analysis and literature review of the genus *Drosophila* and closely related genera (Diptera, Drosophilidae). Insect Syst. Evol. 39, 241–267.

van der Linde, K., Houle, D., Spicer, G.S., Steppan, S., 2010. A supermatrix-based molecular phylogeny of the family Drosophilidae. Genet. Res., Camb. 92, 25–38.

Wang, B.C., Park, J., Watabe, H.A., Gao, J.J., Xiangyu, J.G., Aotsuka, T., Chen, H.W., Zhang, Y.P., 2006. Molecular phylogeny of the *Drosophila virilis* section (Diptera: Drosophilidae) based on mitochondrial and nuclear sequences. Mol. Phylogenet. Evol. 40, 484–500.

Wojtas, K.M., von Kalm, L., Weaver, J.R., Sullivan, D.T., 1992. The evolution of duplicate glyceraldehyde-3-phosphate dehydrogenase genes in *Drosophila*. Genetics 132, 789–797.

Xia, X., Lemey, P., 2009. Assessing substitution saturation with DAMBE. In: Lemey, P. (Ed.), The Phylogenetic Handbook. Cambridge University Press, Cambridge, U.K., pp. 611–626.

Xia, X., Xie, Z., 2001. DAMBE: software package for data analysis in molecular biology and evolution. J. Hered. 92, 371–373.

Xia, X., Xie, Z., Salemi, M., Chen, L., Wang, Y., 2003. An index of substitution saturation and its application. Mol. Phylogenet. Evol. 26, 1–7.

Zachos, J., Pagani, M., Sloan, L., Thomas, E., Billups, K., 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. Science 292, 686–693.

Zhang, W.X., 1993a. Three new species of *nigricolor* species-group of *Drosophila* (*Lordiphosa*) (Diptera: Drosophilidae). Acta Zootaxon. Sin. 18, 220–224.

Zhang, W.X., 1993b. A review of the taxonomic status of the *Lordiphosa denticeps* group with descriptions of four new species (Diptera: Drosophilidae). Entomotaxonomia 15, 144–154.

Zhang, W.X., Liang, X., 1992. Seven new species of the subgenus *Lordiphosa* of *Drosophila*. Acta Zootaxon. Sin. 17, 473–482.

Zhang, W.X., Liang, X., 1994. Three new species of drosophilid flies, Diptera: Drosophilidae, from Yunnan and Hubei, China. Entomotaxonomia 16, 213–219.