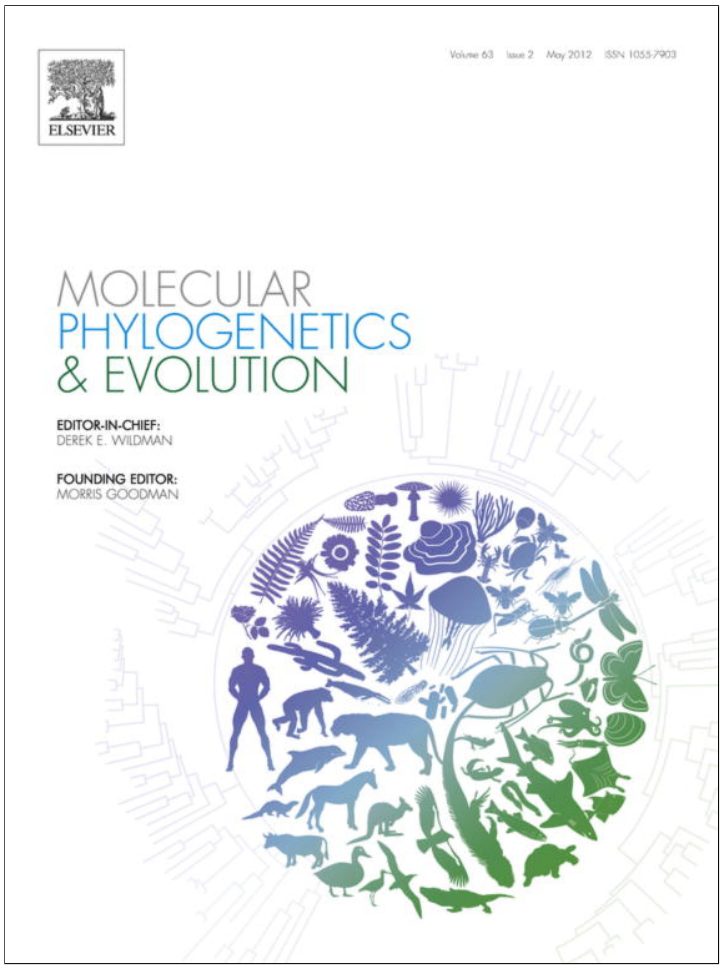

Molecular Phylogenetics and Evolution 63 (2012) 466–474

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

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

Ecological and historical drivers of diversification in the fly genus *Chiastocheta* Pokorny

Anahí Espíndola [a,*], Sven Buerki [b], Nadir Alvarez [c]

[a] Laboratory of Evolutionary Entomology, Institute of Biology, University of Neuchâtel, Emile-Argand 11, 2000 Neuchâtel, Switzerland
[b] Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, United Kingdom
[c] Department of Ecology and Evolution, University of Lausanne, Biophore Building, 1015 Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 28 October 2011
Revised 19 January 2012
Accepted 23 January 2012
Available online 8 February 2012

Keywords:
Chiastocheta
Diversification
Trollius
Biogeography
Phylogeny
Temporal inference

ABSTRACT

Coevolution is among the main forces shaping the biodiversity on Earth. In Eurasia, one of the best-known plant–insect interactions showing highly coevolved features involves the fly genus *Chiastocheta* and its host-plant *Trollius*. Although this system has been widely studied from an ecological point of view, the phylogenetic relationships and biogeographic history of the flies have remained little investigated.

In this integrative study, we aim to test the monophyly of the five *Chiastocheta* eco-morphological groups, defined by Pellmyr in 1992, by inferring a mitochondrial phylogeny. We further apply a new approach to assess the effect of (i) different molecular substitution rates and (ii) phylogenetic uncertainty on the inference of the spatio-temporal evolution of the group.

From a taxonomic point of view, we demonstrate that only two of Pellmyr's groups (rotundiventris and dentifera) are phylogenetically supported, the other species appearing para- or polyphyletic. We also identify the position of *C. lophota*, which was not included in previous surveys.

From a spatio-temporal perspective, we show that the genus arose during the Pliocene in Europe. Our results also indicate that at least four large-scale dispersal events are required to explain the current distribution of *Chiastocheta*. Moreover, each dispersal to or from Asia is associated with a host-shift and seems to correspond to an increase in speciation rates. Finally, we highlight the correlation between diversification and climatic fluctuations, which indicate that the cycles of global cooling over the last million years had an influence on the radiation of the group.

1. Introduction

Coevolution has shaped the diversity of life on Earth. Examples include the evolution of cellular organelles (Margulis, 1970) and the key role of biological interactions in the maintenance of complex ecosystems (Aizen and Ezcurra, 1998; Shanahan et al., 2001; Thompson, 2009). Besides sustaining the diversity on Earth, interactions can also lead to species diversification, especially through adaptive and coevolutionary processes. Cases of diversification driven by coevolution can be found in systems involving specific and strict plant-insect interactions, among which are figs and fig wasps (Cruaud et al., 2011; Wiebes, 1979) and *Yucca* and *Yucca* moths (Pellmyr, 2003; Pellmyr and Huth, 1994). One of the only mutualistic and obligate plant-insect interactions occurring in the Palearctic is that of the nursery pollination system involving species of the globeflower genus *Trollius* L. (Ranunculaceae; 18 species identified in the last revision, Pellmyr, 1992) and their fly pollinators/seed-predators belonging to the genus *Chiastocheta* Pokorny (Diptera: Anthomyiidae).

In the *Trollius*–*Chiastocheta* interaction, each fly species is associated with one or several plant taxa. While adults visit and pollinate flowers and lay eggs on the carpels, larvae are strict seed parasites. Whereas eleven *Chiastocheta* species are known from the adult stage, six additional groups are distributed in restricted Asian zones and have been described by Pellmyr (1992; Table 1) based on the egg's shape and color, and the position where eggs are laid on the carpel.

A first taxonomic treatment of *Chiastocheta* was carried out by Collin (1954), followed by Hennig (1976) and Michelsen (1985). Cladistic inferences based on morphological (egg-shape and color) and behavioral (position of egg-laying) characters conducted by Pellmyr (1992) allowed the delimitation of five major species groups (Table 1). Biogeographically, seven species from all groups are European and involved in nursery pollination mutualistic interactions with *T. europaeus*, the only globeflower species found in the West-Palearctic region (Table 1). Compared to their European relatives, all fly species from the East-Palearctic region are associated

* Corresponding author. Present address: Department of Ecology and Evolution, University of Lausanne, Biophore Building, 1015 Lausanne, Switzerland. Fax: +41 21 692 4165.
E-mail address: mariaanahi.espindola@unil.ch (A. Espíndola).

1055-7903/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2012.01.018

Table 1
Species of *Chiastocheta* investigated in this study. The classification follows Pellmyr (1992), and clades refer to new findings presented here. The geographic range of each species is shown. Species names in bold correspond to poly- or paraphyletic taxonomic entities.

Pellmyr (1992) group	Species/taxa included in the group	Geographic range	Phylogenetic status of the group	Clade (this study)
Rotundiventris	***C. rotundiventris***	Europe	Paraphyletic	I
	R3	Asia		I
	R5	Asia		I
Dentifera	*C. dentifera*	Europe	Monophyletic	II
Trollii	*C. trollii*	Europe	Paraphyletic	III
Inermella	***C. inermella***	Europe	Polyphyletic	III + IV + V
	C. latispinigera	Asia		
	C. lophota	Europe		
	C. pellmyrii	Japan		
Setifera	***C. setifera***	Europe	Polyphyletic	III + V
	C. curvibasis	Asia		
	C. macropyga	Europe		

with several Asiatic *Trollius* species, with which they are less tightly ecologically associated (Pellmyr, 1992).

Previously, Després et al. (2002) used DNA sequences to investigate the phylogenetic status of the five groups of *Chiastocheta* species defined by Pellmyr (1992). Their phylogenetic analyses failed to recover monophyly in three of the five groups and suggested polyphyly of several "species" (see below). Although some taxa were not included in their analyses (*e.g.*, *C. lophota*), causing putative bias in phylogenetic relationships (Funk, 1999), Després et al. (2002) demonstrated that all Asiatic (*C. curvibasis*, *C. latispinigera* and *C. pellmyri*) and three European species (*C. rotundiventris*, *C. macropyga* and *C. dentifera*) were monophyletic. The remaining species (*C. inermella*, *C. trollii* and *C. setifera*) were either para- or polyphyletic, probably due to interspecific gene flow or incomplete lineage sorting of ancestral polymorphism. However, the survey of Després and colleagues lacked extensive sampling, which might have allowed more robust conclusions on the nature of the species, especially considering their large distribution range. Finally, even though dating analyses were performed and biogeographical hypotheses discussed, no quantitative or model-based methods were applied.

Although Pellmyr (1992) and Després et al. (2002) greatly contributed to our understanding of this biotic interaction, their phylogenetic studies were mostly descriptive and did not include multivariate parametric analyses. In this study, we infer the spatio-temporal evolution of the genus *Chiastocheta* using state-of-the-art methods in phylogenetic inference based on a Bayesian approach that allows (i) simultaneous inference of phylogenetic relationships and divergence time between taxa and also (ii) investigation of the effect of different molecular substitution rates on the tempo of speciation and biogeographic scenario of this genus. Moreover, we perform parametric biogeographic analyses by applying model-based biogeographic reconstructions on a set of trees from the Bayesian inference to assess the effect of phylogenetic and dating uncertainty on the biogeographic scenario. The integrative approach proposed here formally evaluates the strength of the inferred spatio-temporal framework and therefore allows discussion of the biogeographic patterns that are independent of phylogenetic and dating uncertainty.

The primary aim of this work is to construct and analyze a complete molecular dataset of all described *Chiastocheta* species for which adults were available [including all species groups in Pellmyr (1992) and the species not included in Després et al. (2002)]. This will clarify the phylogenetic relationships and boundaries of each group and species. Based on this phylogeny, we further infer the spatio-temporal history of the genus and discuss its interaction with *Trollius*.

2. Material and methods

2.1. Sampling

All described *Chiastocheta* species (47 samples total) were obtained from either the field or from genetic databases (*i.e.*, GenBank; Appendix A; see below). Sample selection was done in such a way as to include several specimens of each species, while also sampling from different geographic regions across a species range (to account for any potential intraspecific sequence polymorphism). Following Després et al. (2002), *Delia brassica* L. (Diptera: Anthomyiidae) was used as the outgroup. Field samples were obtained during spring and summer from 2006 to 2008 for all fly species associated with *T. europaeus* and were preserved in 70% ethanol after collection. All flies collected in the field were identified following Hennig (1976) and further confirmed by expert knowledge (V. Michelsen). Sequences of all species not associated with *T. europaeus* were obtained from GenBank (Appendix A).

2.2. Molecular analyses

DNA was extracted from insect legs using the QIAGEN DNeasy Animal tissue extraction kit (QIAGEN, Hombrechtikon, Switzerland) following the manufacturer protocol. The remaining insect parts (abdomen, rest of thorax and head) were preserved in case of further need to confirm species identification.

Two mitochondrial regions (Cytochrome Oxidase I, COI; Cytochrome Oxidase II, COII) were amplified and sequenced. PCR amplification was performed using primers shown in Table 2. Reactions were done in a 20 μl mix, containing 0.5× buffer, between 1 and 2.5 mM $MgCl_2$, 10 mM dNTPs, 1 unit of GoTaq DNA polymerase (Promega, Dübendorf, Switzerland), 0.5 μM primers, 3 μl DNA and run in a TGradient Thermocycler (Biometra, Goettingen, Germany). The program started with 90 s at 95 °C, followed by 40 cycles of 35 s at 95 °C, 1 min at 52 °C, 45 s at 72 °C, and finished with a final elongation of 8 min at 72 °C. PCR products were sequenced by Macrogen Inc. (South Korea). ChromasPro 1.41 (Technelysium Pty. Ltd.) was used to assemble complementary strands and verify software base-calling. New sequences, together with sequences from Després et al. (2002) (Appendix A), were aligned using the CLUSTAL W algorithm (Thompson et al., 1994), as implemented in BioEdit 7.0.4.1 (Hall, 1999). Alignments were manually corrected using BioEdit 7.0.4.1. The presence of NUMTs, and number of constant, variable and parsimony informative base pairs were evaluated using MEGA 4.0 (Tamura et al., 2007).

Possible sequence saturation was examined by calculating the transition:transversion ratios and correlating these values to the genetic distances obtained using the best model of evolution for each region (see below). This was done for each amplified region using DAMBE 5.2.63 (Xia and Xie, 2001).

2.3. Phylogenetic inference and divergence time estimation

To investigate the two main goals of this study – (i) test the validity of the current classification in light of molecular phylogenetic reconstructions and (ii) infer the spatio-temporal evolution of the genus – two sets of phylogenetic inferences were performed. First, maximum likelihood (ML) and Bayesian Markov Chain Monte Carlo (MCMC) approaches were performed to investigate the species boundaries and the monophyly of the five groups as defined by Pellmyr (1992). Second, a coalescence-based Bayesian inference (using different molecular substitution rates; see below) was used to infer the temporal history of the genus *Chiastocheta*.

ML searches were done using RAxML 7.2.8 (Stamatakis, 2006), running 10,000 rapid bootstraps followed by the search of the

Table 2
Regions, primer sequences, annealing temperatures and references used for PCR amplification of COI and COII in *Chiastocheta* species.

Region	Primer	Sequence	Annealing	Reference
COI	COI-2171	TTG ATT TTT TGG TCA YCC NGA AGT	52	Després and Jaeger (1999)
	tRNALeu-3048	TGG AGC TTA AAT CCA TTG CAC		Després and Jaeger (1999)
COII	tRNALeu-3023	GAT TAG TGC AAT GGA TTT AGC TC	52	Després and Jaeger (1999)
	COII-3683	CCR CAA ATT TCT GAA CAT TGA CC		Després and Jaeger (1999)

best-scoring ML tree in one single run and applying the GTRCAT molecular model. A partitioned Bayesian MCMC analysis was done with MrBayes v3.1.2 (Ronquist and Huelsenbeck, 2003). Three cold and six hot chains were run for 50×10^6 generations in two independent runs. The best-fit model (HKY + G for both partitions) was inferred with MrAIC (Nylander, 2004) and defined based on the Akaike's information criterion (AIC; Akaike, 1973). Convergence was accepted when standard deviations reached values lower than 0.01 and when the Potential Scale Reduction Factor index (Gelman and Rubin, 1992) approached 1.0. We considered the MCMC sampling sufficient when the Effective Sampling Size (ESS) was higher than 200 – using TRACER v1.4 (Rambaut and Drummond, 2004). A 50% majority-rule consensus tree was then reconstructed after applying a burn-in of 5×10^6 generations (thus considering a total of 270,006 trees).

In order to infer a temporal framework for the evolution of *Chiastocheta*, we used the partitioned coalescent Bayesian inference approach implemented in BEAST 1.5.3 (Drummond and Rambaut, 2007). The same best-fit models for COI and COII calculated using MrAIC were used in this analysis. Partitions were unlinked for the model of evolution, whereas they were linked for the estimation of the molecular clock and the topology (see below). Two runs of 30×10^6 generations were done, sampling one tree every 1000 generations. Due to the lack of fossils for the genus *Chiastocheta*, direct calibration of the topology was not possible and divergence times were therefore estimated by applying gene-specific mtDNA substitution rates. According to Papadopoulou et al. (2010), the two most extreme insect mitochondrial molecular clocks based on relaxed methods and fossil data correspond to means of 0.9% My^{-1} (0.78–1.02% My^{-1}; Zakharov et al., 2004) and 2.28% My^{-1} (1.66–2.9% My^{-1}; Wahlberg, 2006). We favoured an exploratory approach considering these two drastically different substitution rates independently, to investigate the effect of this parameter on divergence time estimation as well as on the biogeographic history of *Chiastocheta* (see below). For each of the two substitution rate values, plus one that took into account their mean value (1.59% My^{-1}), we ran a partitioned coalescent Bayesian analysis following Borer et al. (2010). In the three cases, a relaxed clock with log-normal branch length distribution was used and a Yule speciation model was applied to model population size through time (other prior parameters were set to default; Drummond et al., 2006). For each parameter, convergence of runs was confirmed by the examination of their respective distributions in TRACER 1.4 (Rambaut and Drummond, 2004). After removing a 10% burn-in period (thus considering a total of 54,002 trees), a maximum clade credibility tree with median branch lengths was built using TreeAnnotator 1.5.4 (Drummond and Rambaut, 2007). Because all approaches provided the same topology, we kept only the maximum credibility clade tree (with median branch lengths) from the mean value (1.59% My^{-1}) and displayed the 90% interval of confidence on node ages by using the 5% and 95% percentiles of the 0.9% My^{-1} and the 2.28% My^{-1} time estimations, respectively. This was done using a collection of R scripts (R Development Core Team, 2012) available on request to SB, and allowed us to present both the phylogenetic estimation and the temporal uncertainty related to the consideration of different molecular substitution rates on a unique topology.

2.4. Biogeographic analyses

Following recommendations made by Buerki et al. (2011), areas were mainly defined based on the geological criterion (sensu Sanmartín and Ronquist, 2004) together with the current species distributions. Three areas were thus considered: (i) Asia (defined as all the lands east of the Ural Mountains); (ii) Europe (including all the western Eurasian territory and extending up to the Ural Mountains) and (iii) Japan. Terminals were coded as in Espíndola et al. (2010), according to the location of each sample.

The dispersal–extinction–cladogenesis (DEC) likelihood model implemented in Lagrange v.2.0.1 (Ree et al., 2005; Ree and Smith, 2008) was used to investigate the biogeographic history of the *Chiastocheta* species. This method is a parametric, extended version of the dispersal–vicariance analysis (Ronquist, 1997) that estimates ancestral ranges, transition rates between ranges, and biogeographical scenarios of range inheritance for a group of taxa in a maximum likelihood framework, taking tree branch-lengths into consideration (Ree and Sanmartín, 2009). Another advantage of the DEC model is its ability to adapt a transition matrix (*i.e.*, *Q* matrix) to reflect the changing paleogeography, connections between areas (*e.g.*, land bridges) through time, or dispersal capabilities of the group of interest (Buerki et al., 2011). By taking advantage of this flexibility, the *Q* matrix for analyses performed in this study was constrained according to dispersal possibilities between areas (Appendix B). A dispersal probability of 1.0 was thus set when areas were directly connected, whereas – in order to maintain the reducibility of the Markov Chain (Buerki et al., 2011) – it was defined to 0.01 instead of 0 when areas were not connected. To take phylogenetic and dating uncertainty into account while inferring the biogeographic scenario of *Chiastocheta*, the Lagrange analysis was run on 300 randomly selected trees from the BEAST analyses (*i.e.*, 100 trees from each substitution rate) and ancestral areas on nodes were summarized on the 1.59% My^{-1} maximum credibility clade tree. Following the approach described in Buerki et al. (2011), ancestral area reconstructions for each node were plotted on the tree using pie charts. To investigate the effect of dating and phylogenetic uncertainty on the inferred biogeographic scenario, a 3D plot was built for each node depicting (i) the ancestral area reconstructions, (ii) the probability of assignment to a given area and (iii) the node age. This chart is fundamental in proving the strength of the ancestral area reconstruction, when taking the dating uncertainty into consideration. The collection of R scripts (R Development Core Team, 2012) to achieve this task are available on request to SB.

A problem often encountered in biogeographic analyses is the effect of unbalanced taxon sampling that interferes with biogeographic reconstructions (Ree and Sanmartín, 2009). In the present study, taxa were sampled in a way to minimize this bias, even if all regions were not similarly represented because they do not harbor the same number of species.

The number of lineages diversifying through time and the effect of climate on the diversification pattern observed in the data were analyzed using the package *ape* (Paradis et al., 2004) and a set of R (R Development Core Team, 2012) scripts. We constructed a Lineages Through Time (LTT) plot (for the three maximum credibility clade trees) using several *ape* (Paradis et al., 2004) commands, and

we further combined the plots with the climatic registers inferred for the last 3.5 My. These registers are the relative values of isotopic Oxygen ($\delta^{18}O$, expressed in ‰) calculated by Zachos et al. (2001) from deep-sea foraminifera calcite, which track the evolution of deep-sea temperature and of continental ice volumes. An increase in these $\delta^{18}O$ values corresponds to a decrease in temperatures, while a decrease indicates climate warming (Zachos et al., 2001). In order to directly quantify a possible correlation between the number of lineage splits and climatic variations, we defined 100,000 year periods and calculated for each of them (i) the number of new lineages, (ii) the maximum $\delta^{18}O$ value and (iii) its variation from period to period (Delta $\delta^{18}O$). Afterwards, we adjusted linear models between (i) and (ii), and (i) and (iii), using the *lm* function in R (R Development Core Team, 2012). This protocol was applied to the phylogenetic inferences obtained using the three different rates of evolution.

2.5. Host-plant evaluation

Because the host-plant might have influenced the evolution and diversification pattern of the seed predator/pollinators, we plotted the distribution of host-plants on the BEAST maximum credibility clade tree. A relationship between host-plant shifts and speciation events would then suggest that the exploitation of a new resource might trigger diversification. The information of the exploited host-plant for each fly species was obtained from the global survey done by Pellmyr (1992).

3. Results

3.1. Phylogenetic relationships

The final alignment comprised 1136 bp (COI: 657 bp; COII: 479 bp) and contained no gaps and no NUMTs. Between 25% (COI) and 44% (COII) of the positions were variable (Table 3). None of the amplified regions were saturated when considering the ingroup only, while some saturation was observed when including the outgroup for the region COI (Appendix C).

Table 3
Number and percentage of constant (C), variable (V), parsimony informative (PI) and total base pairs (Length) for the *Chiastocheta* datasets of COI and COII. Percentages are shown between parentheses.

Mitochondrial region	C	V	PI	Length (base pairs)
COI	489 (74.43)	168 (25.57)	77 (25.57)	657
COII	268 (55.95)	211 (44.05)	39 (8.14)	479

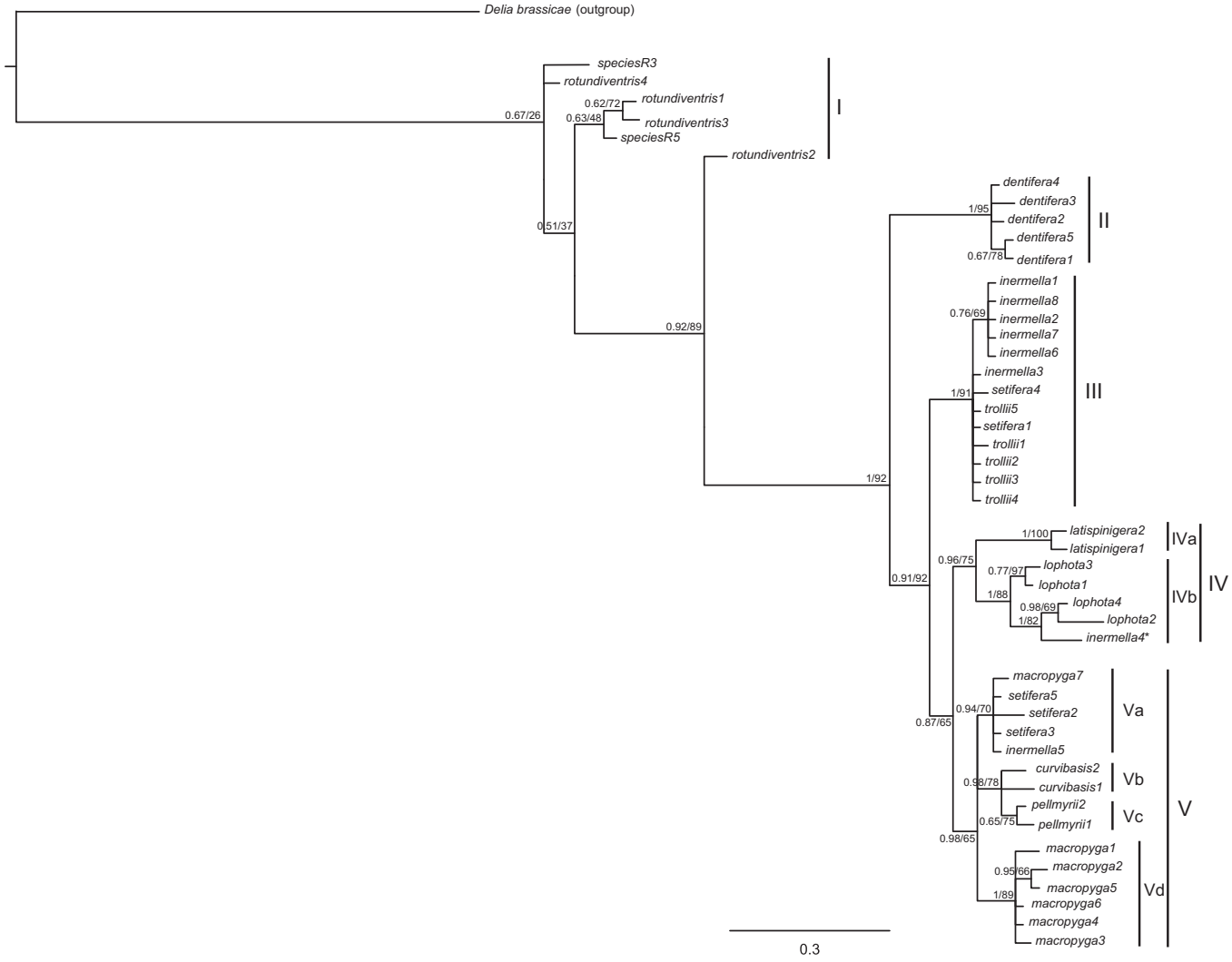

Fig. 1. Bayesian 50% majority-rule consensus tree of the *Chiastocheta* genus. Bayesian posterior probability and ML bootstrap values are displayed on branches. The star indicates a putative misidentified sample in a previous study (Després et al., 2002).

While MrBayes and RAxML topologies were congruent (Fig. 1), an incongruence was observed with the BEAST topology. This involved the clade comprising *C. dentifera* (Figs. 1 and 2). Because this incongruence does not influence our conclusions, we will use the first two approaches (*i.e.*, RAxML and MrBayes) to discuss the phylogenetic relationships, while we will use the second (*i.e.*, BEAST) to investigate the spatio-temporal evolution of the genus.

Our results identified five groups and six subgroups (Fig. 1), supporting two of the groups previously proposed by Pellmyr (Table 1) and identifying several para- and polyphyletic taxa. The most basal lineages forming a grade (grade I) correspond to the paraphyletic rotundiventris group (species delimitation within this group is still obscure and we have referred to other taxa as R3 and R5, following Pellmyr, 1992). Clade II contained all the accessions belonging to the dentifera group (which comprises only a single species, *C. dentifera*). Clade III included all the trollii group samples, all *C. inermella* and some *C. setifera*. Clade IV included *C. latispinigera* (subclade IVa) and *C. lophota* (subclade IVb), thus partially corresponding to the previously defined inermella group. Finally, clade V corresponded to the setifera group and included also *C. pellmyrii* (described as belonging to the inermella group).

Five species – *C. dentifera* (clade II), *C. latispinigera* (subclade IVa), *C. lophota* (subclade IVb), *C. curvibasis* (subclade Vb) and *C. pellmyri* (subclade Vc) – were monophyletic in our phylogenetic inferences, whereas the remaining species appeared to be either para- or polyphyletic. The case of *C. macropyga* is of note, since most of its representatives cluster in a well-supported subclade (Vd), but one specimen is included in another subclade (Va).

3.2. Divergence time estimations and biogeographic history

Based on the 1.59% My^{-1} maximum credibility clade tree from the BEAST analysis (displaying the 90% confidence interval on nodes according to the other substitution rates), the origin of the genus is estimated to the Middle Pliocene [ca. 3.36 million years ago (Mya)] (Fig. 2). The two main clades diverged at the Pliocene–Pleistocene boundary around 2.72 Mya, and most of the diversification events happened from 1.66 Mya to the present (Calabrian and Holocene).

When the difference in temporal estimations and phylogenetic uncertainty are taken into account, the biogeographic estimations were consistent for all nodes (data not shown), besides the one at the base of *C. rotundiventris* (node 50 in Fig. 3a), which was more frequently assigned to Asia at older ages, and otherwise assigned to Asia–Europe (Fig. 3b).

The ancestral area reconstruction (Fig. 3a) inferred a European origin of the genus, during the Middle Pliocene (Placenzian). The biogeographic scenario inferred four dispersal events: three from Europe to Asia between 1.5 and 1 Mya (clades I, IV and V), and

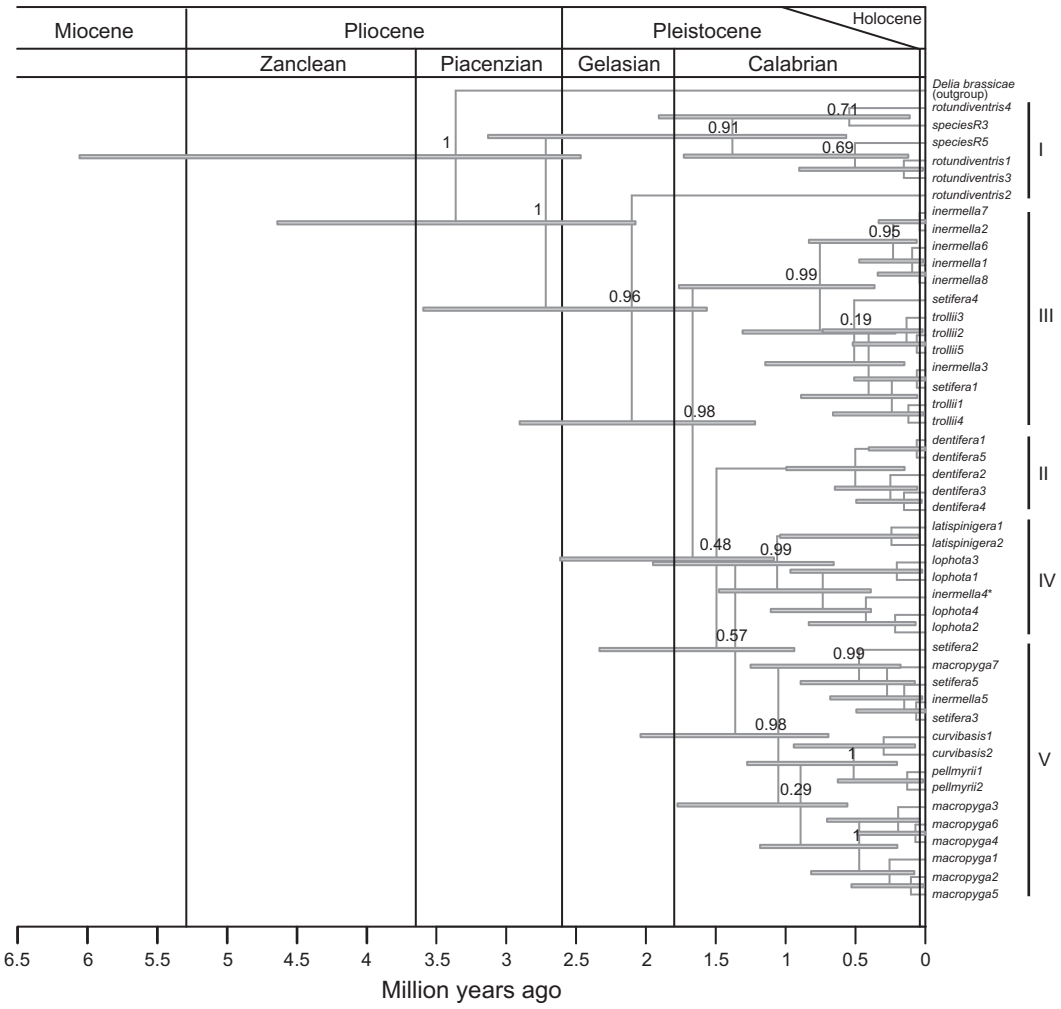

Fig. 2. Bayesian maximum credibility tree of *Chiastocheta* with median branch length inferred from the middle substitution rate (1.59% My^{-1}). Bayesian posterior probabilities are shown on nodes. Bars indicate errors on time estimations, considering the 5% and 95% percentiles of distribution of the ages obtained from the minimum (0.9% My^{-1}) and maximum (2.28% My^{-1}) substitution rates, respectively. Clades are shown, as in Fig. 1. The star indicates a putative misidentified sample in a previous study (Després et al., 2002).

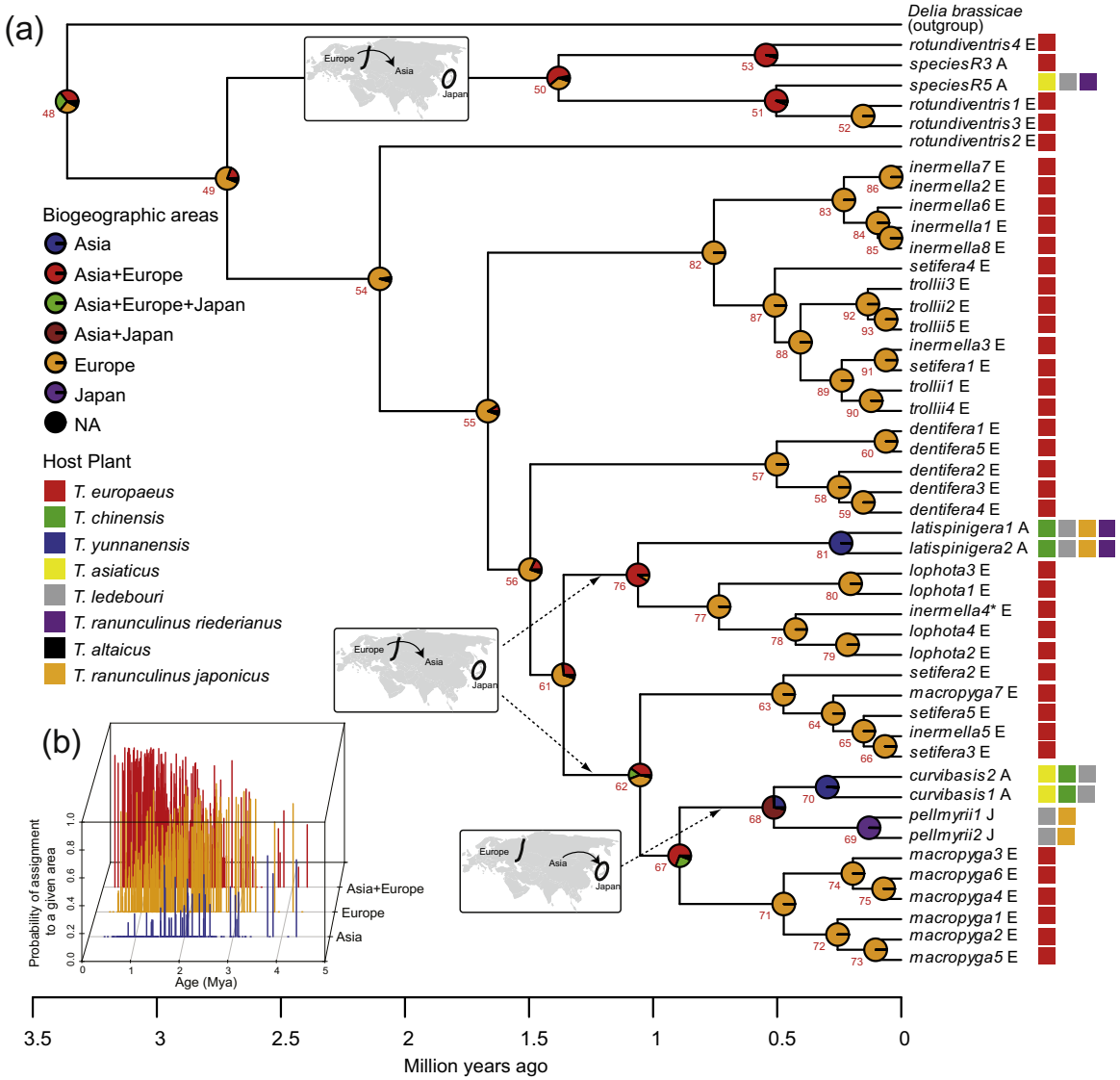

Fig. 3. (a) DEC biogeographic scenario inferred on 300 randomly selected trees from the different substitution rates and displayed on the Bayesian maximum credibility tree. Values on nodes indicate node number. Pie-charts correspond to ancestral area(s) probabilities for each node, with colors indicating ancestral areas. Colors in squares right to the tips indicate host-plant(s) of the respective species. The star indicates a putative misidentified sample in a previous study (Després et al., 2002). Letters following taxon names correspond to biogeographic areas (A: Asia, E: Europe, J: Japan). (b) Effect of the dating uncertainty on the biogeographic reconstruction for the node at the basis of the rotundiventris group (node 50 in (a)). Bars represent the probability of assignment (*y* axis) to a given area (*z* axis), taking the age of the node into account (*x* axis).

one from Asia to Japan occurring between 0.9 and 0.5 Mya (clade V).

The ecological relationship with *T. europaeus* is phylogenetically widespread in the genus, and the exploitation of different host-plants occurred in clades I, IV and V (Fig. 3a). By combining biogeographic and host-plant evaluations, it appeared that every dispersal event led to a host-plant shift and to further speciation.

The analysis of the variation of the number of lineage splits in relation with climatic oscillations indicated that from the Quaternary onwards, the number of phylogenetic splits drastically increased in the genus *Chiastocheta* (Fig. 4a). The results obtained using the three different mutation rates identified a significant correlation between the number of lineages and the $\delta^{18}O$ variance from period to period (Fig. 4b; 1.59% My^{-1}: slope = 3.811, $t_{(33)} = 3.901$, $p < 0.001$; $R^2 = 0.32$, $F_{(1,33)} = 15.22$, $p < 0.001$; other rates: data not shown), indicating that the climate can explain part of the diversification in this genus. Our results also demonstrated a significant correlation between the maximum $\delta^{18}O$ values for each period and the number of new Lineages Through Time (Fig. 4b; 1.59% My^{-1}: slope = 2.8983, $t_{(33)} = 3.330$, $p < 0.01$; $R^2 = 0.25$, $F_{(1,33)} = 11.09$, $p < 0.01$; other rates: data not shown), indicating that diversification was facilitated by shifts towards lower temperatures.

4. Discussion

4.1. Phylogenetic inference of the genus Chiastocheta

Compared to the previous work of Després et al. (2002), our results suggest different among-clade relationships (clades III, IV and V), as well as a new delimitation of some species (*e.g.*, *C. lophota*) (Figs. 1–3; Table 1). The topological differences between these studies might be due to our inclusion of all currently described species, as well as of several accessions per species (representative of the morphological and geographic ranges within species). Our analyses thus complete and update the inferences already obtained by previous studies.

We cannot confirm the phylogenetic position of all groups previously defined based on morphological and behavioral characters (Pellmyr, 1992). First, the rotundiventris group (grade I) and *C.*

rotundiventris are paraphyletic in our study (Fig. 1). However, because the grade appears to be strongly different and isolated of the remaining taxa, we consider that the rotundiventris group should be validated. Adding more DNA regions in future studies would probably help clarifying this situation.

Clade II includes all *C. dentifera*. Additionally, this group is characterized by several distinctive morphological and behavioral characters (*e.g.*, a smooth egg surface and the fact that the species lays eggs between carpels; see Pellmyr, 1992 for more details), which might be considered as diagnostic in future taxonomic revisions.

The three remaining clades (clades III, IV and V) appear much more complicated. First, none of these clades fully fit any of the groups as defined by Pellmyr (1992). Second, some species occur in more than one single clade (*e.g.*, *C. setifera*). Taken together, this might suggest that taxonomic revisions are critically needed. The phylogenetic framework presented here is therefore a valuable asset for describing new taxonomic groups and identifying synapomorphies. Another possibility is hybridization between taxa, which should also be considered if defining new taxonomic groups.

Clade III corresponds mostly to the trollii group – it contains all the *C. trollii* specimens included in this study – and also includes half of the *C. setifera* and most *C. inermella* samples. In the case of the latter species, a subclade was formed by *C. inermella* samples from Scandinavia, suggesting the occurrence of spatial genetic structure potentially leading to local differentiation (Figs. 1–3). However, this pattern has to be confirmed by expanding the sampling.

Clade IV – including *C. latispinigera* and *C. lophota* – was not recovered in Després et al. (2002) and is the result of the inclusion of samples of *C. lophota* (Fig. 1). We should note that considering the substantial genital morphological similarities (Dr. Michelsen, pers. com.), it is highly likely that the sample identified as C. *inermella* (*inermella4*) by Després et al. (2002) belongs to *C. lophota*. Considering our results, *C. latispinigera* and *C. lophota* are sister species, likely to have diverged as the consequence of a dispersion event from Europe to Asia (see below, Fig. 3a).

Finally, clade V only partially corresponds to the setifera group because it also contains *C. pellmyri*. Here, some subclades match species delimitations: Vb, *C. curvibasis*; Vc, *C. pellmyri* (Fig. 1). It is possible that although *C. macropyga* is relatively well defined genetically (subclade Vd), some hybridization is occurring with other taxa (with *C. setifera*, for example). An alternative explanation might be that the taxonomy of this species requires revision. The phylogenetic position of *C. setifera* is difficult to resolve, since it is not restricted to any clade or subclade, and, moreover, it is strongly polyphyletic (it occurs in clades III and V). It is likely that

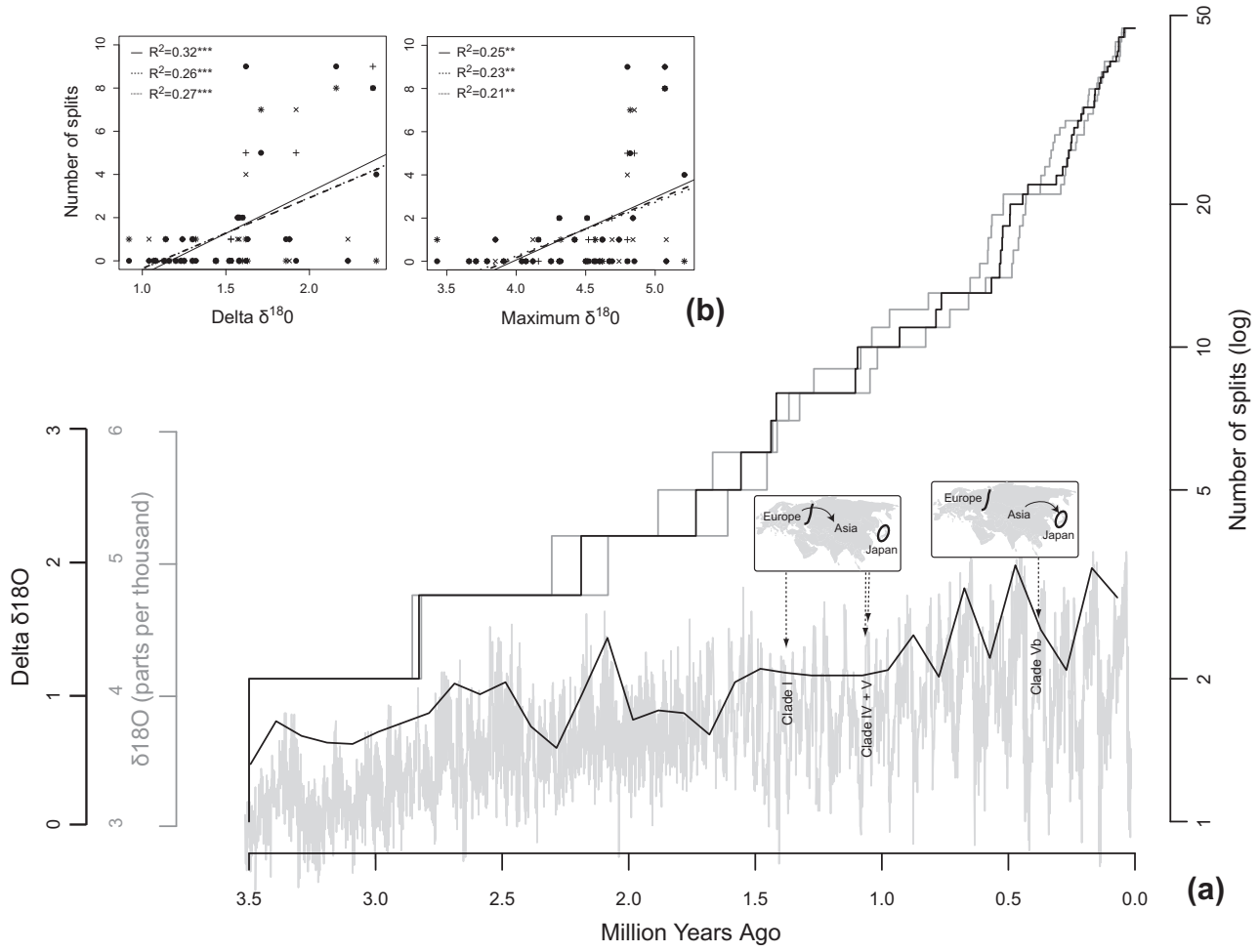

Fig. 4. (a) Evaluation of the number of splits in the phylogeny, as a function of time, and in relation with the climatic record from the last 3.5 My. Climatic oscillations are shown in gray in the background (obtained from Zachos et al. (2001)), with variance shown in black (scales are colored in the same way than curves). The timing of the number of splits obtained for the middle substitution rate is shown with a black line; those corresponding to the minimum (0.9% My^{-1}) and maximum (2.28% My^{-1}) substitution rates are shown with gray lines. Arrows indicate the timing of the dispersion events shown in the captions, and the concerned clades are indicated. (b) Correlations were calculated between the number of splits registered for each 100,000 years period and the variation of $\delta^{18}O$ from a period of 100,000 years to the previous, or the maximum $\delta^{18}O$ value obtained for each period. Correlations, R^2 and significance levels are shown for inferences corresponding to the three different rates of molecular evolution: black circles and straight line: mean value (1.59% My^{-1}); greek cross and hatched line: minimum value (0.9% My^{-1}); saltire and dotted line: maximum value (2.28% My^{-1}). Significance levels: $^{***}p$-value < 0.001; $^{**}p$-value < 0.01.

this taxon represents two different species, or that it is able to hybridize with species belonging to different clades (*e.g.*, *C. trollii*, *C. inermella*, *C. macropyga*). To understand the complex patterns appearing in this clade, further analyses should investigate nuclear regions, since only maternally mitochondrial inherited markers have been used here. Furthermore, additional analyses are also needed to define morphological synapomorphies supporting the phylogenetic clades and describe species.

4.2. Spatio-temporal evolution of Chiastocheta: influence of climate change on the distribution of taxa and host-plant shifts

Recent studies (Papadopoulou et al., 2010) have pointed out the importance of using relaxed molecular clocks, as well as considering substitution rate heterogeneities. Using this approach, our divergence time estimations indicate that the group diverged recently (Fig. 2), with the first main split in the genus happening close to the Pliocene–Pleistocene boundary (2.71 Mya; and potentially going back to 4.63 Mya). These results differ from those obtained by Després et al. (2002), who applied a fixed molecular clock and used a standard substitution rate of 2% My^{-1} (Brower, 1994), to reach an estimation of the origin of the genus not exceeding 2 Mya.

Besides providing a new temporal framework for understanding the history of the genus and the initiation and establishment of coevolutionary processes, our findings suggest that the strong climatic cooling represented by the Pliocene–Pleistocene boundary (currently placed at 2.6 Mya, Walker and Geissman, 2009) and the high frequency and amplitude of climatic oscillations characterizing the last million years have had an influence on the spatial history and diversification events of the genus *Chiastocheta*. Our results demonstrate that the global shift into cooler climates favored an increase in the number of lineages of this cold-adapted genus, probably guided by range expansions and the later occupation of new ecological niches (*e.g.*, host-plants; Fig. 3a). Indeed, it has been demonstrated in other cold-adapted species (Stewart et al., 2010; Triponez et al., 2011) that, during cold periods, the distribution range of these organisms was extended and, in such a situation, dispersion was favored. It is particularly striking that the divergence of the genus, including its oldest split and much of the latter diversification events, happened in periods in which strong shifts into colder conditions happened. This same pattern was also recovered by our correlative analyses, which demonstrated that in the case of *Chiastocheta*, climatic variations leading to cooler conditions were associated with diversification events (Fig. 4b).

Compared to other biogeographic studies, our survey applied a biogeographic approach that allows inclusion of the effects of phylogenetic and dating uncertainty in the reconstruction of ancestral areas. This multifaceted integrative approach demonstrated that ancestral node area reconstructions were in general independent of the dating and phylogenetic uncertainty. An exception to this rule was node 50 (Fig. 3b). However, although the inferred ancestral areas differed for this node, this did not affect the main biogeographic conclusions. Nevertheless, this example points out the importance of widely exploring the effect of dating and phylogenetic uncertainty on biogeographic inferences, which should be evaluated in future biogeographic studies.

Our biogeographic inferences (Fig. 3a) indicate that only four large-scale dispersal events are required to explain the current distribution of *Chiastocheta* species. First, the most recent common ancestor of *C. rotundiventris* dispersed from Europe to Asia sometime between the Pliocene–Pleistocene boundary and the Calabrian age (2.71–1.38 Mya). Later (during the Calabrian age; 1.36–1.05 Mya), while a European lineage survived in Europe, two lineages dispersed to Asia. Finally, from there, one of these lineages dispersed to Japan between 0.89 and 0.51 Mya (Calabrian age). As mentioned above, these results also indicate that lineages diversified and dispersed during the Pleistocene, a period recognized by its marked glacial ages (Ehlers and Gibbard, 2004). In the northern hemisphere, the distribution of biomes was shifted to southern regions during these periods, and arctic species appeared to be much more widespread than today (*e.g.*, Brochmann et al., 2003). It is possible that these distributional expansions facilitated dispersal of flies of this genus into new ecological conditions, allowing interactions on new host-plants and further specialization. This might explain the current presence of some species of the rotundiventris group in isolated areas such as southern China (Pellmyr, 1992). The colonization of Japan might have also been facilitated by glacial periods, during which the Japan Sea's level was around 130 m lower than today (Nakada et al., 1991), creating a land bridge linking the Archipelago to South Korea for at least 2 My (Kameda and Kato, 2011).

Our results suggest that the origin of this plant–insect interaction is European, with several switches to new host-plants happening when fly lineages dispersed to Asia and came into contact with other already established *Trollius* species occurring in the plant genus' center of diversification (Després et al., 2003). This scenario is compatible with the distribution of the host-plant use in the phylogeny (Fig. 3a). Indeed, the exploitation of *T. europaeus* is phylogenetically basal and each dispersion event is followed by the adaptation to new host-plants, leading to resource specialization and further speciation processes. Under the coevolutionary and biogeographic scenario presented here, *T. europaeus* is the plant species that might have been coexisting with *Chiastocheta* flies for most of the time, probably favoring the selection of the strong morphological, ecological, behavioural and physiological coadaptations observed today (*e.g.*, Collin, 1954; Gallet et al., 2007; Pellmyr, 1989; Pompanon et al., 2006). From a strict coevolutionary point of view, this might also explain the fact that only the European plant species (*T. europaeus*) is currently harboring a strict nursery mutualism with the flies (Pellmyr, 1992). From another perspective, while all *Chiastocheta* feed on *Trollius*, the opposite is not true, suggesting, as mentioned by Després et al. (2002), that cospeciation between the plant and the insect did not happen, and that while *Trollius* diversified in Asia (Després et al., 2003), *Chiastocheta* only lately invaded the area and specialized on some *Trollius* species.

5. Conclusion

In the present study, we demonstrated that relationships among species within the genus *Chiastocheta* are different to initial predictions. Based on our results, it is evident that group definitions have to be improved and species redefined. We recommend that in the meantime, ecological studies on the coevolutionary dynamics of these insects do not ignore that genetic boundaries of all typological species are not as strict as formerly supposed. Indeed, this could have a non-negligible influence in the interpretation of ecological and coevolutionary surveys.

The biogeographic approach presented here, which relies on different evolutionary rates, consistently recovered ancestral areas and biogeographic events for virtually all nodes. It thus appears that even if the timing of these events can slightly vary, their position in the phylogenetic inference is constant. The method used here is therefore useful when analyzing the effect of the phylogenetic and dating uncertainty on the biogeographic history of a set of taxa.

Acknowledgments

The authors would like to thank the field and technical assistance of Y. Triponez, R. Lavigne, N. Magrou, D. Gyurova,

P. Lazarevic, H. Hipperson, T. Suchán, N. Russel, P. Bowler, R. Daffydd, M. Howe and V. Michelsen. They also thank Sarah M. Hird for checking the English. This project was funded by the Swiss National Science Foundation (Project No. 3100A0-116778). Financial support to SB was provided by a Marie-Curie Intra-European Fellowship (CRADLE; No. 253866).

Appendices A–C. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2012.01.018.

References

Aizen, M.A., Ezcurra, C., 1998. High incidence of plant–animal mutualisms in the woody flora of the temperate forest of southern South America: biogeographical origin and present ecological significance. Ecol. Austral 8, 217–236.
Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Second International Symposium on Information Theory, Kiado, A., Budapest, pp. 267–281.
Borer, M., Alvarez, N., Buerki, S., Margraf, N., Rahier, M., Naisbit, R.E., 2010. The phylogeography of an alpine leaf beetle: divergence within *Oreina elongata* spans several ice ages. Mol. Phylogen. Evol. 57 (2), 703–709.
Brochmann, C., Gabrielsen, T.M., Nordal, I., Landvik, J.Y., Elven, R., 2003. Glacial survival or *tabula rasa*? The history of North Atlantic biota revisited. Taxon 52, 417–450.
Brower, A.V.Z., 1994. Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. Proc. Natl. Acad. Sci. USA 91, 6491–6495.
Buerki, S., Forest, F., Alvarez, N., Nylander, J.A.A., Arrigo, N., Sanmartín, I., 2011. An evaluation of new parsimony-based versus parametric inference methods in biogeography: a case study using the globally distributed plant family Sapindaceae. J. Biogeogr. 38 (3), 531–550.
Collin, J.E., 1954. The genus *Chiastocheta* Pokorny (Diptera: Anthomyiidae). Proc. R. Entomol. Soc. London (B) 23, 95–102.
Cruaud, A., Cook, J., Da-Rong, Y., Genson, G., Jabbour-Zahab, R., Kjellberg, F., Pereira, R.A.S., Rønsted, N., Santos-Mattos, O., Savolainen, V., Ubaidillah, R., van Noort, S., Yan-Qiong, P., Rasplus, J.Y., 2011. Fig–fig wasp mutualism: the fall of the strict cospeciation paradigm. In: Patiny, S. (Ed.), Evolution of Plant–Pollinator Interactions. Cambridge University Press, London, UK, pp. 68–102.
Després, L., Jaeger, N., 1999. Evolution of oviposition strategies and speciation in the globeflower flies *Chiastocheta* spp. (Anthomyiidae). J. Evol. Biol. 12, 822–831.
Després, L., Pettex, E., Plaisance, V., Pompanon, F., 2002. Speciation in the globeflower fly *Chiastocheta* spp. (Diptera: Anthomyiidae) in relation to host plant species, biogeography, and morphology. Mol. Phylogen. Evol. 22 (2), 258–268.
Després, L., Gielly, L., Redoutet, W., Taberlet, P., 2003. Using AFLP to resolve phylogenetic relationships in a morphologically diversified plant species complex when nuclear and chloroplast sequences fail to reveal variability. Mol. Phylogen. Evol. 27 (2), 185–196.
Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.
Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4 (5), e88.
Ehlers, J., Gibbard, P.L., 2004. Quaternary glaciations: extent and chronology. In: Developments in Quaternary Science. Elsevier, Amsterdam, San Diego.
Espíndola, A., Buerki, S., Küpfer, P., Bedalov, M., Alvarez, N., 2010. New insights into the phylogeny and biogeography of *Arum* L. (Araceae): unravelling its evolutionary history. Bot. J. Linn. Soc. 163, 14–32.
Funk, D.J., 1999. Molecular systematics of cytochrome oxidase I and 16S from *Neochlamisus* leaf beetles and the importance of sampling. Mol. Biol. Evol. 16 (1), 67–82.
Gallet, C., Ibanez, S., Zinger, L., Taravel, F.R., Trierweiler, M., Jeacomine, I., Després, L., 2007. Plant chemical defense induced by a seed-eating pollinator mutualist. J. Chem. Ecol. 33 (11), 2078–2089.
Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. 7, 457–472.
Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.
Hennig, W. (Ed.), 1976. Anthomyiidae. Die Fliegen der Palaearktischen Region. Stuttgart, E. Schweizerbart.
Kameda, Y., Kato, M., 2011. Terrestrial invasion of pomatiopsid gastropods in the heavy-snow region of the Japanese Archipelago. BMC Evol. Biol. 11, 118.
Margulis, L., 1970. Origin of Eukaryotic Cells. Yale University Press, New Haven, Connecticut, USA.
Michelsen, V., 1985. A revision of the Anthomyiidae (Diptera) described by J.W. Zetterstedt. Steenstrupia 11 (2), 37–65.
Nakada, M., Yonekura, N., Lambeck, K., 1991. Late Pleistocene and Holocene sea-level changes in Japan: implications for tectonic histories and mantle rheology. Palaeogeogr. Palaeoclimatol. Palaeoecol. 85, 107–122.
Nylander, J.A.A., 2004. Mraic.pl. Uppsala, Sweden, Distributed by the Author.
Papadopoulou, A., Anastasiou, I., Vogler, A.P., 2010. Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. Mol. Biol. Evol. 27 (7), 1659–1672.
Paradis, E., Claude, J., Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290.
Pellmyr, O., 1989. The cost of mutualism – interactions between *Trollius europaeus* and its pollinating parasites. Oecologia 78 (1), 53–59.
Pellmyr, O., 1992. The phylogeny of a mutualism: evolution and coadaptation between *Trollius* and its seed-parasitic pollinators. Biol. J. Linn. Soc. 47, 337–365.
Pellmyr, O., 2003. Yuccas, Yucca moths, and coevolution: a review. Ann. Missouri Botan. Garden 90 (1), 35–55.
Pellmyr, O., Huth, C.J., 1994. Evolutionary stability of mutualism between yuccas and yucca moths. Nature 372, 257–260.
Pompanon, F., Pettex, E., Després, L., 2006. Patterns of resource exploitation in four coexisting globeflower fly species (*Chiastocheta* sp.). Acta Oecol. 29 (2), 233–240.
R Development Core Team, 2012. R: A Language and Environment for Statistical Computing.
Rambaut, A., Drummond, A.J., 2004. Tracer v1.4. Edinburgh.
Ree, R.H., Sanmartín, I., 2009. Prospects and challenges for parametric models in historical biogeographical inference. J. Biogeogr. 36, 1211–1220.
Ree, R.H., Smith, S.A., 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Syst. Biol. 57, 4–14.
Ree, R.H., Moore, B.R., Webb, C.O., Donoghue, M.J., 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. Evolution 59, 2299–2311.
Ronquist, F., 1997. Dispersal–vicariance analysis: a new approach to the quantification of historical biogeography. Syst. Biol. 46, 195–203.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19 (12), 1572–1574.
Sanmartín, I., Ronquist, F., 2004. Southern hemisphere biogeography inferred by event-based models: plant versus animal patterns. Syst. Biol. 53 (2), 216–243.
Shanahan, M., So, S., Compton, S.G., Corlett, R., 2001. Fig-eating by vertebrate frugivores: a global review. Biol. Rev. Camb. Philos. Soc. 76 (4), 529–572.
Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22 (21), 2688–2690.
Stewart, J.R., Lister, A.M., Barnes, I., Dalén, L., 2010. Refugia revisited: individualistic responses of species in space and time. Proc. R. Soc. B 277, 661–671.
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.
Thompson, J.N., 2009. The coevolving web of life. Am. Nat. 173 (2), 125–140.
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 (22), 4673–4680.
Triponez, Y., Buerki, S., Borer, M., Naisbit, R.E., Rahier, M., Alvarez, N., 2011. Discordances between phylogenetic and morphological patterns in alpine leaf beetles attest to an intricate biogeographic history of lineages in postglacial Europe. Mol. Ecol. 20 (11), 2442–2463.
Wahlberg, N., 2006. That awkward age for butterflies: insights from the age of the butterfly subfamily Nymphalinae (Lepidoptera: Nymphalidae). Syst. Biol. 55, 703–714.
Walker, J.D., Geissman, J.W., 2009. Geologic Time Scale: Geological Society of America. The Geological Society of America.
Wiebes, J.T., 1979. Co-evolution of figs and their insect pollinators. Annu. Rev. Ecol. Syst. 10, 1–10.
Xia, X., Xie, Z., 2001. DAMBE: software package for data analysis in molecular biology and evolution. J. Hered. 92 (4), 371–373.
Zachos, J., Pagani, M., Sloan, L., Thomas, E., Billups, K., 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. Science 292 (5517), 686–693.
Zakharov, E.V., Caterino, M.S., Sperling, F.A.H., 2004. Molecular phylogeny, historical biogeography, and divergence time estimates for swallowtail butterflies of the genus *Papilio* (Lepidoptera: Papilionidae). Syst. Biol. 53 (2), 193–215.