UNIVERSITÉ DE NEUCHÂTEL

Inferring reciprocal evolutionary histories in associated species of plants and insects in two European pollination systems

Thèse présentée à la Faculté des Sciences
Institut de Biologie
Université de Neuchâtel

Pour l'obtention du grade de docteur ès sciences
par

María Anahí Espíndola

Acceptée sur proposition du jury:

Dr. Nadir Alvarez, co-directeur de thèse
Prof. Martine Rahier, co-directeur de thèse
Prof. Bryan C. Carstens (Louisiana State University, USA), rapporteur
Dr. Laurence Després (Université de Grenoble, France), rapporteur
Dr. Marc Gibernau (CNRS Toulouse, France), rapporteur
Dr. Felix Gugerli (WSL, Birmensdorf, Zürich, Suisse), rapporteur
Prof. Philippe Küpfer (Université de Neuchâtel, Suisse), rapporteur

Soutenue le 17 Novembre 2010

Université de Neuchâtel
2010

Université de Neuchâtel

FACULTE DES SCIENCES
Secrétariat-Décanat de la faculté
Rue Emile-Argand 11
CH-2000 Neuchâtel

IMPRIMATUR POUR LA THESE

Inferring reciprocal evolutionary histories in associated species of plants and insects in two European pollination systems

María Anahí Espíndola

UNIVERSITE DE NEUCHATEL

FACULTE DES SCIENCES

La Faculté des sciences de l'Université de Neuchâtel,
sur le rapport des membres du jury

Dr Nadir Alvarez (co-directeur de thèse, Université de Neuchâtel),
Prof. Martine Rahier (co-directrice de thèse, Université de Neuchâtel),
Prof. Bryan C. Carstens (Louisiana State University USA),
Prof. Laurence Després (Université de Grenoble F),
Dr Marc Gibernau (CNRS Toulouse F),
Dr Felix Gugerli (WSL, Birmensdorf CH),
Prof. Philippe Küpfer (Université de Neuchâtel)

autorise l'impression de la présente thèse.

Neuchâtel, le 22 novembre 2010

Le doyen :
P. Kropf

Téléphone : +41 32 718 21 00 E-mail : secretariat.sciences@unine.ch www.unine.ch/sciences

«Toy with different perspectives. Look for the unusual. Try consciously to innovate. Train yourself to imagine new schemes and innovative ways to fit the pieces together. Seek the joy of discovery. Always test your new thoughts against the facts, of course, in rigorous, cold-blooded, unemotional scientific manner. But play the great game of the visionary and the innovator as well.»

J. E. Oliver

« Gracias a la vida, que me ha dado tanto… »

V. Parra

A Cristina y Ricardo, mis padres.

Contents

Summary / Résumé..…… ix

Acknowledgements/ Remerciements…….…….. xiii

General introduction ……..…… 1

Chapter one ……..…… 33
New insights into the phylogenetics and biogeography of Arum (Araceae): unravelling its evolutionary history

Chapter two ……..…… 55
Molecular relationships in the sub-family Psychodinae (Diptera: Psychodidae)

Chapter three ……..……..… 85
Variation in the proportion of flower visitors of Arum maculatum along its distributional range in relation with community-based climatic niche analyses

Chapter four ……..…… 99
Size-advantage model and pollination strategies in plants

Chapter five ……..…….. 131
Comparative phylogeography in a specific and obligate pollination antagonism

Chapter six ……..…… 163
Evolutionary history of Chiastocheta species interacting with Trollius europaeus

Chapter seven ……..…….. 209
No fossils, no answers? Inferring postglacial genetic consequences in cold-adapted species when neither ancient DNA nor macroremains are available

Chapter eight..…… 247
Phylogeography of mutualists and the effect of the host on the genetic structure of its partners

General Discussion ……..... 283

General Conclusion ……..…….. 295

Appendix 1: *Curriculum vitae* ……..……..... 299

Keywords: specific interactions, coevolution, *Arum maculatum*, *Trollius europaeus*, *Chiastocheta*, Psychodidae, antagonism, mutualism, ecological niche model, comparative phylogeography, biogeography, diversification, hindcasting methods, arctic-alpine species, Size-Advantage Model, taxonomy, statistical phylogeography

Summary

Coevolution is defined as reciprocal evolutionary changes that might arise at any spatiotemporal scale. Despite every organism on Earth undergoes coevolutionary interactions, cases of one-to-one specific relationships are generally rare. However, because of the reduced number of interacting species they concern, these species-specific associations are interesting to evolutionary biologists because they allow testing hypotheses in simple frameworks.

Despite the history and evolution of coevolutionary interactions have been studied in several cases in the last decade, this topic remains difficult to fully circumscribe because of the multiplicity of factors that affect one or the other species concerned. Moreover, a lot is known about coevolution at a small scale, but little has been done at larger and more integrative scales spanning wider spatiotemporal ranges.

Phylogeography is a young area of biology that allows understanding the distribution of lineages in space and time. Despite that the idea of parallely studying the history of species involved in specific interactions appears simple, this has rarely been done until now probably because of the technical efforts this would represent.

From a theoretical point of view, we could propose that in specific interactions, because of the dependence between the partners involved, we should observe some phylogeographic pattern associated to the type of interaction studied. In this way, while partners of mutualistic interactions should present similar postglacial histories, this should not be true for those associated by antagonistic relationships.

In this thesis, we exploit different techniques and approaches to test this general hypothesis. The final aim of this study is thus to understand if it is possible to identify a pattern of comparative phylogeography in relation to the type of interaction, using as case-studies two specific and obligate European interactions: the antagonistic relationship established between *Arum maculatum* L. (Araceae) and its Psychodid (Diptera) pollinating flies, and the nursery pollination mutualism involving *Trollius europaeus* L. (Ranunculaceae) and the *Chiastocheta* (Diptera: Anthomyiidae) species complex.

Before testing our comparative phylogeographic hypotheses and because studying the phylogeography of interactions requires a wide knowledge of the environmental, taxonomic and historical frameworks in which these ecological relationships arose, it was first needed to clearly delimitate the identity of species, their distribution and the environmental factors

influencing their survival to finally understand their comparative history. We thus took advantage of the potentialities that interdisciplinary approaches provide, applying molecular taxonomy, biological and evolutionary methods, biogeographic inferences, ecological niche models and hindcasting techniques, as well as classical and recently-developed phylogeographic analyses.

Our results indicate that the phylogeographic patterns of these specific and obligate antagonistic and mutualistic relationships appear to be related to the type of interaction. Antagonistic partners presented incongruent phylogeographic patterns, what can be notably explained by differences in their life-history traits. Species involved in mutualistic interactions partly showed congruent phylogeographic patterns (particularly in the cases of *T. europaeus* and *C. dentifera*). Flies interacting with *T. europaeus* appear moreover to present different histories, regardless of their important ecological similarities.

These results demonstrate that the systems studied appear to be far more complex than initially supposed, with crossed effects of environmental and historical features on the dynamics of the interaction. Because of the high complexity and interdependency of factors affecting one or the other partner, performing investigations in an interdisciplinary framework appears indispensable to disentangle the dynamics of interactions.

Mots-clés: interactions spécifiques, coévolution, *Arum maculatum*, *Trollius europaeus*, *Chiastocheta*, Psychodidae, antagonisme, mutualisme, modèle de niche écologique, phylogéographie comparée, biogéographie, diversification, méthodes de prédictions de distributions passées, espèce arctique-alpine, modèle d'avantage de taille, taxonomie, phylogéographie statistique

Résumé

La coévolution est définie comme la série de changements réciproques apparaissant entre des organismes. Malgré le fait que tous les organismes suivent des processus de coévolution, les cas des relations spécifiques un-un sont relativement rares. Ces types d'interactions spécifiques sont par contre très intéressants pour les évolutionnistes, notamment à cause du nombre réduit d'organismes qu'elles comprennent, ce qui permet de tester des hypothèses évolutives dans un cadre simple.

Malgré le fait que l'histoire et l'évolution des interactions coévolutives sont des processus assez explorés, le thème reste difficile à étudier spécialement à cause du grand nombre de facteurs influençant les espèces concernées. D'un autre côté, beaucoup est à ce jour connu à des petites échelles spatio-temporelles, alors que très peu a été fait à des dimensions plus larges et plus intégrées.

Le jeune domaine de la phylogéographie s'intéresse à la compréhension de la distribution des lignées génétiques dans l'espace et dans le temps. Malgré que l'idée de l'étude parallèle des histories phylogéographiques de plusieurs espèces est un concept simple, il n'a été exploré que pour quelques cas isolés.

D'un point de vue théorique, nous pouvons proposer que dans des interactions spécifiques, nous devrions attendre à voir des patrons phylogéographiques liés au type d'interaction étudiée, spécialement à cause de l'interdépendance des espèces concernées. Alors que des partenaires mutualistes devraient présenter des histories démographiques similaires, ceci ne devrait pas être le cas des espèces liées par des interactions de type antagoniste.

Le but ultime de ce travail est alors de comprendre s'il est possible d'identifier des patrons phylogéographiques spécifiques aux types d'interactions. Afin de répondre à cette question, nous testons nos hypothèses sur deux interactions européennes spécifiques et obligées: l'antagonisme formé par *Arum maculatum* L. (Aracée) et ses pollinisateurs de la famille Psychodidae (Diptères) ; le mutualisme *de crèche* présent entre *Trollius europaeus* L. (Ranunculacée) et le complexe d'espèces du genre *Chiastocheta* (Diptère: Anthomyiidae).

Comme avant de tester nos hypothèses phylogéographiques, il est nécessaire de posséder une connaissance approfondie des cadres environnemental, taxonomique et historique dans lesquels les interactions sont apparues et se développent, nous avons

d'abord défini les limites des espèces en relation, circonscrit leur distribution et compris les facteurs environnementaux influençant la survie des entités, pour ensuite inférer leurs histoires évolutives. Nous avons profité des potentialités que des approches interdisciplinaires fournissent, appliquant des méthodes de taxonomie moléculaire, des approches biologiques et évolutives, des inférences biogéographiques, des méthodes de modélisation de niche écologique et des prédictions des distributions passées, ainsi que des analyses phylogéographiques classiques et modernes.

Nos résultats indiquent que les patrons phylogéographiques des espèces en interaction sont liés aux types de relations étudiées. Ceux des espèces antagonistes sont incongruents et peuvent être expliqués par des différences dans les traits d'histoire de vie des espèces. Les espèces mutualistes ont des patrons phylogéographiques partiellement congruents (comme le cas de *T. europaeus* et *C. dentifera*). Le complexe d'espèces de mouches interagissant avec *T. europaeus* présenterait alors des histoires différentes, malgré ses similarités écologiques importantes.

Nos résultats démontrent que ces systèmes sont beaucoup plus complexes qu'initialement supposés, probablement à cause des effets croisés entre environnement et histoire. A cause de cette complexité et de l'interdépendance des facteurs affectant l'un ou l'autre partenaire, des études interdisciplinaires apparaissent alors comme extrêmement appropriées lorsque notre but est de comprendre des interactions.

Acknowledgements / Remerciements / Agradecimientos / Bedankungen

J'aimerais remercier tout d'abord mon directeur de thèse Nadir Alvarez. C'est grâce à son patience, sa motivation et ses apports que cette thèse a pu être finie. Et c'est surtout grâce à son sens de l'humour que des critiques sur des phrases sorties de ma petite tête du type « among all these incongruent groups, the most congruent are …» ou bien « cytoplasmic impurity » sont devenues des choses plus compréhensibles pour des gens autres que moi-même et aussi –surtout- une source de blagues. Nadir : MERCI ! (comme ça, en grand, comme un grand cri !).

Si Nadir a été une sorte de papa professionnel, c'est Yann qui a été mon frère. C'est durant ces longues journées de terrain, ces heures incomptables de voiture, ces milliers de kilomètres parcourus, ces aventures partagées que je me suis rendue compte que je suis bien chanceuse d'avoir eu un co-thésard comme lui. Si maintenant je parle un peu mieux le français et je ne dis plus (bon d'accord, *presque* plus) *dans* la route, mais *sur* la route, et non pas *caléfaction* mais *chauffage*, c'est aussi complètement de sa faute. Si mon chemin académique est parsemé de personnes comme lui, ce sera sans doute un plaisir de continuer à le parcourir !

I would like to thank all the members of my thesis committee. Their comments and critics have largely improved this work. Each of them in a different way made me aware of points that I have ignored and provided me the occasion to become a better scientist. I thank them for their time, their interest and the energy they spent in leading me in my learning process. I would also like to thank all the European scientists who generously shared their knowledge, and who demonstrated me that science is the art of sharing. A special thanks goes to Bryan Carstens, who hosted me in his lab, and who leaded me through the "long and winding road" of the coalescent theory and statistical phylogeography.

Une thèse ne peut probablement pas voir le jour dans un milieu de travail désagréable. J'aimerais alors remercier à toute l'équipe (présente et passée) e-vol/FARCE, et spécialement à mes « cobureautaires » : Greg, Mättu, Tom et Kevin. Je suis bien consciente de la chance que j'ai eue de vous avoir comme collègues de bureau. Merci pour la bonne humeur, pour le soutien et pour votre patience !

I am lucky of being so lucky. It is almost incredible that I did not only have ONE but TWO places full of great colleagues! The Carstens lab staff is the reason why my time in Louisiana was so fruitful –both from a personal and a professional point of view. It is thanks to all this great people that I could i) learn some of the rules of football, ii) understand some of the mysteries of phylogeny, iii) explore the amazing Louisianan nature, iv) transform my opinion about the US, v) feel like "at home" far away from home.

Un gigantesque merci va aussi pour les différentes personnes avec qui j'ai partagé une partie des nombreux kilomètres faits lors de mon travail de terrain : Pat, Nat, Reine. Merci pour toutes ces aventures et pour ne pas avoir eu peur de prendre la route avec moi !

Comme dans toute procédure académique, commencer, faire et finir une thèse demande une quantité considérable de paperasse. Je ne veux même pas imaginer ce que cela aurait été si Brigitte Cattin et Natacha Schneiter n'étaient pas là ! C'est une chance pour l'Institut d'avoir des secrétaires aussi efficaces et compétentes.

A nivel personal quiero agradecer a mis padres. Creo que sin su amor incondicional y su apoyo no sería lo que soy. Son ellos los que me enseñaron a no tenerle miedo al trabajo, y los que me muestran cada día que la honestidad y el compartir son también maneras de amar. Esta tesis les pertenece tanto como a mí.

Mis padres no sólo me dieron la vida, pero también la suerte de tener un hermano. Gracias hermanito por tu amor y por las aventuras pasadas y por venir!

Mine Familie sind au die Webers. Wenn man weg isch vo Dihei es ist so schön zu wüsse das man au eine Familie hier hat! Merci !

Aussi un eternel merci à Susanne Junod, qui a permis que toute cette aventure commence. Si le monde était plein de gens comme elle, ce serait certainement un endroit nettement plus joyeux et plus juste. Là où elle est, je suis sure qu'elle sourit en me voyant écrire ces lignes.

Merci aux amigos, d'ici et de allá. A los de Santa Fe, de Córdoba, del Sur, a los de México… et à ceux de Neuch' et de Lausanne ! C'est tan lindo de vous avoir ! Vous remplissez mi vida et son mi familia de coeur.

Merci URS. Merci für dine Geduld, dine Liebi, dine ewige Freude und diner Unterstützig. Ich liebe dich, und wenn was kommt wird so schön si als was scho passiert isch, ich kann mich nur vo de Zukunft freude!

General Introduction

1. Coevolution

Since the formal definition of the coevolutionary theory in 1964 (Ehrlich and Raven 1964) numerous studies have demonstrated that coevolutionary processes shape the Earth's biodiversity (Thompson 2009). From the evolution of cellular organelles (Margulis 1970) to the main role played by interacting species in the maintenance of complex environments (*e.g.,* fruit production in the cases of fig and fig-wasps (Shanahan, So et al. 2001); ecosystem survival in Southern temperate forests (Aizen and Ezcurra 1998); the effect of corals in maintaining reef environments (Thompson 2009)), coevolutionary processes appear to have been fundamental for the development of diversity of life.

Coevolution is defined as the reciprocal evolutionary change in interacting species (Thompson 1994) and is generally detected by the presence of coevolved traits appearing between the interacting organisms (Lunau 2004). The theory was first described based on observations of closely related plants which present also closely related specific herbivores (Ehrlich and Raven 1964). It was then surprising that related plants presented specific secondary compounds which helped them to be protected against herbivory, and that herbivores feeding on those plants were also closely related and appeared specifically resistant to the plants' toxins. In order to present an explanation of these unexpected congruences it was proposed that a "chemical escalation" of plant chemical defenses existed, which would have created a way for the plant to escape herbivory, but which would also have liberated new "empty" ecological niches, which could be occupied by "new" resistant herbivores. If the process was repeated several times, every physiological change would have the potential to delimitate a new taxon in both the plant and the herbivore, what was described as the "escape and radiate" process (Thompson 1994).

In fact, it is not random that the theory was first described for plants and insects because in these two groups coevolutionary processes are extremely frequent. Interactions between plants and insects are not only abundant (i.e., at least 400.000 species of herbivorous insects have been described up to now, Mitter, Farrell et al. 1991), but they appear to have

been happening for hundred million years (Farrell, Mitter et al. 1992). Parallel plant and insect fossil analyses have thus suggested that the increase in the diversification of angiosperms in the mid-Cretaceous was almost simultaneous to radiations happening in groups of insects specialized on pollination as Apoidea or Lepidoptera (Grimaldi 1999). As mentioned above, coevolution was initially "discovered" while investigating herbivory relationships, however it is not only restricted to it, and diversification in angiosperms is currently proposed to be especially a consequence of adaptations related to another type of interaction: zoophilous (animal-mediated) pollination (Crepet 1996). It has been demonstrated that mediated pollination directly influenced the rate of diversification of plants through its effect on floral innovations (Crepet 2000).

Considering the tightness of the process, different types of coevolutionary interactions have been proposed, with a separation along a continuum between one-to-one or specialized interactions (*i.e.,* occurring between a reduced number of species) and diffuse coevolution or generalism (*i.e.,* more than two interacting species) (sensu Lunau 2004). Another classification concerns the benefit they provide to the interacting species. While a mutualism is defined as an interaction in which each of the species obtains a net benefit through the "exchange of commodities in a *biological market*" (Bronstein, Dieckmann et al. 2004), in antagonistic interactions one species increases its own fitness by diminishing the other(s)'(s). Despite that these two types of interactions can be either specialized or generalist, it has been observed that most interactions involve far more than two species (Ollerton, Killick et al. 2007). However, in spite of being rare, specific interactions represent extraordinary natural laboratories, in which the forces of coevolution are the most evident and thus the easiest to observe.

Specialization is thought to have arisen from generalist interactions (Futuyma and Moreno 1988). It has been proposed that specialization increases fitness in terms of a raise of the benefit obtained from the interaction thanks to an optimization of the exploitation of the resource, or through the avoidance of interspecific competition (for a review see Pellmyr

2002). However, it has also been shown that not all cases of generalism are ancestral and that under certain situations specific interactions are not energetically positive, what appears to force transitions from specificity to generalism (Tripp and Manos 2008). The potential and capabilities to do this shift appear however to be species-dependent, and while some specific taxa have been able to rapidly change resources (e.g., Vestiaria coccinea, Smith, Freed et al. 1995), others experience fitness reductions after environmental changes induced the absence of their specific partners (e.g., long-term lack of seed production in absence of pollinators, Aizen and Feinsinger 1994).

Despite that these ideas have been around for a while, the number of studies focusing on this topic is limited (Tripp & Manos, 2008). One explanation for the small amount of information available is that most studies addressing questions of adaptive evolution concern changes happening (or having happened) at large temporal scales (*e.g.*, million years). Much less has been done for understanding the stability and coevolutionary aspects of the history of species at shorter temporal scales (*e.g.*, thousand years). At this scale, however, it could be possible to see the effect of differential demographic and distributional changes of the species involved, and their influence on the maintenance or loss of the interaction, especially in relation with major environmental changes.

2. The coevolutionary process in an ecological niche and spatial framework

Ecological niches have been initially defined by Hutchinson (1957) and include all conditions that allow a species to survive. These conditions include both abiotic (*e.g.,* climatic, topographic, edaphic) and biotic (*i.e.,* inter-specific interactions) factors. The ecological niche of a species thus defines not only its “role” in the community in which it is included; it also has a direct effect on the capability of survival of the species at a given location. A species will thus survive only if conditions defined by its ecological niche are realized at specific locations.

Because environmental conditions are included in the definition of the ecological niche of a species, any variation in the abiotic factors has the potential to directly affect the distribution range of organisms. Accordingly, environmental variations have been demonstrated to influence the distributions of species. This fact has been shown to apply both for past (e.g., Cheddadi, Vendramin et al. 2006) as well as for current climatic changes (e.g., Walther, Post et al. 2002).

Similarly, the same effect is expected for the biotic component of the ecological niches, especially if the interaction clearly increases the fitness of the organism (Pellissier, Alvarez et al. in press). In a case where the presence of the associated species defines the survival of another species, the probability of survival and maintenance of the dependent species will be tightly related with the presence of its associated taxa. Consequently, distribution ranges of interacting species must overlap, at least partially, and if this is not the case, another species should fulfill the role of the one absent; otherwise the organism will not be able to survive at a given location.

The niches of all organisms are influenced both by biotic and abiotic factors. However, in species participating in obligate and specific interactions, the effect of the biotic component in the definition of the realized niche is primordial, since a reduced number of factors (*i.e.*, presence or absence of particular species) can drastically define the output of the equation allowing survival. In these species, because the biotic dependence relies on a restricted number of partners, their absence can drastically endanger the maintenance of the dependent species.

3. Phylogeography, or the effect of abiotic factors on the history of species

In comparison to other areas of biology, phylogeography is a relatively new domain in evolutionary biology interested in the study of the spatial variation of diversity at the intraspecific level (Avise 2009). The aim of this area of research is to identify the

geographical distribution of genetic lineages, considering that phylogenies are not only influenced by time but also by space.

During the last decades, numerous studies have investigated the phylogeography of species (e.g., Hewitt 2001; Schmitt 2007). Based on these studies, it could be demonstrated that climatic variations related with the Quaternary glaciations (2.6 Million years ago - present) had a non negligible effect on the spatial distribution of the genetic diversity of species (Hewitt 2000; Taberlet and Cheddadi 2002). In fact, it has been shown that the cyclic nature of these series of glacial and interglacial periods induced repeated organismal range shifts, which followed the distribution of their suitable habitat (e.g., Velichko, Novenko et al. 2007).

Unlike southern areas of the globe, the European subcontinent presents several characteristics that make it interesting when studying phylogeographic patterns. First, it is found at relatively high latitudes (higher than 45-40° latitude North) and presents an important exposed surface. Glacial events have thus an effect on the European area since polar ice-sheets easily reach the continent, and the impact of climatic variations is therefore extremely direct. Second, the region is bordered in the South by an internal sea, the Mediterranean, which acts as a climatic buffer, bringing heat and humidity, especially during cold periods, but blocking southward dispersal of organisms. Finally, the continent presents at its southern range a series of peninsulas (Iberia, Italy, Balkans), separated to the north of the rest of the continent by mountain ranges (Pyrenees, Alps, Carpathians-Dinaric mountains). During cold and dry periods as those identified during glacial events, these features create conditions in southern regions that are extremely favorable to temperate species, because they harbor warmer and moister conditions.

Geological studies concurred in demonstrating that during the last cold global events (Last Glacial Maximum, LGM, 21-18 thousand years ago –Kya), northern ice-sheets advanced in Europe as far south as northern Poland and Germany and covered the British Isles. At the same time, more southern mountain chains featured an extension of its glaciers

(Ehlers and Gibbard 2004). Moreover, because of the concentration of water in ice-sheets the sea level fell and several regions appeared interconnected by dry land due to the emergence of landbridges in the English Channel and the Adriatic. The presence of ice caps induced not only climatic variations (lower temperatures, drier climate), but also variations in the distribution of habitats on the remaining exposed land (Hewitt 1999). Considering this, it has been demonstrated that the tundra limit –today restricted to northern Europe and mountain zones– was shifted to central Europe, and that permafrost was present in regions as far south as southern France and northern Bulgaria (Hewitt 1999). These large environmental variations influenced organisms, which had to follow the cline of their suitable habitat by changing their distributions (*i.e.*, their realized ecological niche was spatially displaced). It has thus been proposed that, because conditions were milder in the southern peninsulas, temperate species found refugia in one or several of these regions (Stewart, Lister et al. 2010).

Taking these hypotheses into account, phylogeographic studies investigate the spatial distribution of the genetic diversity in different species distributed throughout the European area. These researches demonstrated that in temperate species there was a genetic signature indicating that southern regions harbored more important levels of genetic diversity than northern zones, less adapted for survival during glacial periods (Taberlet, Fumagalli et al. 1998). This pattern was explained by the fact that during glacial periods, different preglacial populations would have moved to these southern refugial areas, carrying with them their preglacial genetic diversity. From this point of view, each refugium would have thus potentially harbored different preglacial genotypes. Once climate warmed up, recolonization of ice-released lands would have happened from these refugia, with only some individuals invading successfully the deglaciated areas, and inducing an associated reduction in genetic diversity in those areas, a process called “leptokurtic” recolonization (Hewitt 1996). The consequence of this dynamics is that if we currently sampled the whole distribution ranges of temperate species, we would observe regions of high diversity, which

are generally placed in southern zones, with a drastic reduction in diversity when moving to the north.

Because the phylogeography of European organisms has been widely studied, it is today possible to identify in this continent regions of high and low genetic diversities (whatever species is considered), which are interpreted as refugia and postglacial recolonized areas, respectively (e.g., Taberlet, Fumagalli et al. 1998; Schmitt 2007). An integrated analysis of these studies allowed the identification and proposition of several phylogeographic patterns for the west-Palearctic, which are currently considered as phylogeographic paradigms.

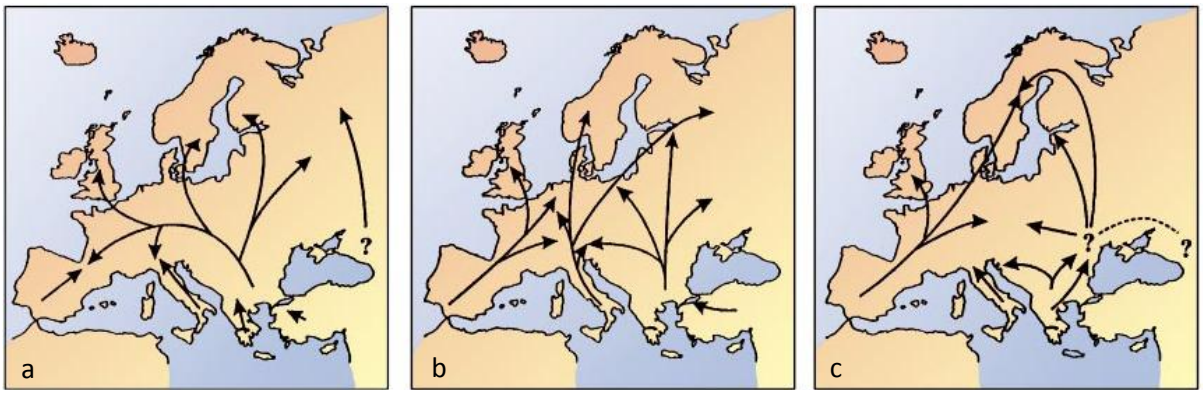


Figure 1 – Phylographic paradigms of post-glacial recolonization of Europe (Hewitt 2000). a – the grasshopper (*Chorthippus parallelus*); b – the hedgehog (*Erinaceus europeus/concolor*); c – the bear (*Ursus arctos*).

These paradigms are illustrated by the three species in which they were initially discovered (Figure 1): i) the grasshopper *Chorthippus parallelus*, which found shelter in three (or four) southern refugia, among which only the Balkanic lineage has recolonized the current distributional range; ii) the hedgehog *Erinaceus europeus*/*concolor* that found refugia in the three southern European peninsulas, from each of which recolonization of the current range happened; iii) the bear *Ursus arctos*, which was restricted to three (or four) refugia, with recolonization of the current area happening from the Iberian and the Balkanic/Carpathian refugia (Hewitt 2000).

Recently, palynological and genetic surveys have also informed that regions occupied by temperate species during glacial periods were not necessarily restricted to the southern peninsulas, but were also present in more northern latitudes, as in some regions south of the Carpathians or in north-western France (Stewart and Lister 2001; Provan and Bennett

2008). It is thus possible that, while main refugia were restricted to southern areas, other would have been present in more northern locations that harbored locally milder conditions than those observed in surrounding areas (regions recognized today as "cryptic refugia", Stewart, Lister et al. 2010).

Contrasting with the large number of studies centered on temperate species, far less is known about phylogeographic patterns of cold-related species (Brochmann, Gabrielsen et al. 2003). Probably the most important difficulty in the understanding of phylogeographic patterns for these species is that, unlike temperate species, they are currently experiencing populational contractions, and that regions they currently occupy should be considered as refugia (Stewart, Lister et al. 2010). Moreover, these species are generally difficult to sample, occupying mountainous areas or high latitudes regions.

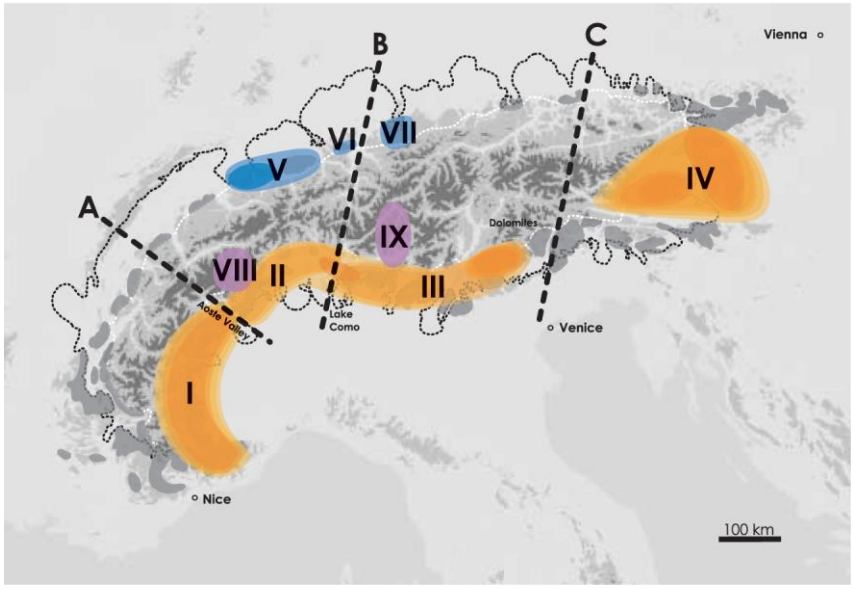


Figure 2 – Glacial refugial zones of mountain plant species in the Alps (Schönswetter, Stehlik et al. 2005). Colors indicate type of refugia: southern, central and northern. Broken lines indicate borders of biogeographic zones.

Despite that several arctic-alpine species have been already studied in Europe, their phylogenies either informed on the absence of clear genetic structure in the region (e.g., Alvarez, Thiel-Egenter et al. 2009) or on high admixture levels in southern mountain ranges (e.g., Schönswetter, Tribsch et al. 2004; Ehrich, Gaudeul et al. 2007). Moreover, the only area for which a phylogeographic paradigm similar to those of temperate species was proposed is the Alps (Figure 2; Schönswetter, Stehlik et al. 2005). In this region, the presence of genetic groups specific to the east, center, south and west of the mountain

range was identified in several plant species, with a clear differentiation in the distribution of refugial regions guided by soil chemistry (Alvarez, Thiel-Egenter et al. 2009).

Despite the fact that phylogeographic patterns of cold-related species are less investigated than those of temperate species, it is currently largely accepted that phylogeographic patterns in Europe are well known. However, little is known about the effect of ecological interactions on these phylogeographic patterns. Even though, as mentioned above, in the same way that climatic variations are able to force and guide the distribution of the suited habitat of a species, biotic interactions could also potentially influence the distribution of species through their effect on the definition of the realized niche.

4. Comparative phylogeography of interacting organisms, or the effect of biotic factors on the fate of species

Although the role played by biotic factors on the fitness of interacting species has been known for a long time (for a review, see Bronstein, Dieckmann et al. 2004), its effect on the intra-specific phylogeography of species has been rarely tested (e.g., Tsai and Manos 2010). In the current dissertation, I propose to test a simple hypothesis postulating that the type of interaction considered (mutualistic *vs.* antagonistic) influences the role that biotic factors have on the history of species. More precisely, because in mutualistic interactions partners depend one on the other, we could hypothesize that the survival and fate of each species should follow congruent phylogeographic patterns (*i.e.*, shared refugia and recolonization pathways). In contrast, in antagonistic interactions, where only one species depends on the presence of the other, we would expect a lack of congruence in phylogeographic patterns, since the independent species is able to disperse and move “freely”, while the dependent species systematically needs the presence of the other to assure its survival.

Trying to investigate these questions is the main topic of my PhD. During the time of my investigations, I have thus worked on two different biological models consisting of

emblematic European plant-insect pollination interactions, respectively mutualistic and antagonistic.

5. The models

Specific and obligate interspecific interactions are ideal models to test hypotheses related with the effect of the interaction on the level of congruence of phylogeographic histories of ecologically associated organisms. However, probably because of the important number of adaptations required for a specific interaction to evolve and maintain, and considering the risks associated with relying on only one partner for the survival of the species (especially in a variable abiotic environment), the number of these types of interactions is relatively low, particularly outside the tropics.

In the Palearctic, several of these interactions have been identified in systems involving plants and insects. Despite that they do not represent one-to-one interactions, they include a small number of interacting species, and in each case, the survival of a dependent species is threatened if the other species is not present.

5a. A case study of a mutualistic interaction – Trollius europaeus L. and Chiastocheta spp. Pokorny

Nursery pollination systems represent a special case of entomophilous mutualistic pollination: insects visit the flowers and actively or passively pollinate it, laying eggs on the developing flowers/fruits since larvae feed specifically on the plant reproductive organs (Dufaÿ and Anstett 2003). Extremely specialized examples of this system are fig-fig wasps (Jousselin, van Noort et al. 2008), yucca-yucca moths (Pellmyr and Huth 1994), *Silene latifolia-Hadena bicruris* moths (Kephart, Reynolds et al. 2006) or *Trollius-Chiastocheta* flies (Pellmyr 1989). Among these, only the latter biological system has been shown to be widespread in the Palearctic region.

The genus *Trollius* (Ranunculaceae) is composed of between 18 and 31 geophyte species depending on the author (Doroszewska 1974; Pellmyr 1992; Després, Gielly et al.

2003). It presents a distribution associated to cold meadows and shows a range occupying the whole Palearctic kingdom. Because of its ecological requirements, it is distributed in lowlands at Northern latitudes higher than 50°, south of which it is present only in mountainous areas (alpine environments). Some species are also partially present in restricted regions of the Nearctic (*i.e.*, *T. riederianus, T. laxus* and *T. ledebouri*). A special feature that makes this genus interesting is that most species are visited and pollinated by different species of the Anthomyiid flies genus *Chiastocheta* Pokorny (Pellmyr 1992).

Flies of the genus *Chiastocheta* are exclusively distributed in the Palearctic, and it has been proposed that their centre of origin is eastern Asia (Pellmyr 1992). Posterior molecular studies suggested, however, that the distribution of the genus was first restricted to the European region and that only afterwards they invaded Eastern areas (Després, Pettex et al. 2002). Up to now, between 10 and 18 species have been described, based on different morphological characters (Hennig 1976; Michelsen 1985; Pellmyr 1992). All species are extremely specialized on the genus *Trollius*, whose seeds are eaten by the fly larvae (Pellmyr 1992), with no records of these species laying eggs on plants not belonging to the genus *Trollius.*

Among *Trollius* species, *T. europaeus* L. obtained the attention of biologists because of several special features. First, from a biogeographical point of view, this species is the only present at the Western edge of the Palearctic (Figure 3a), being isolated from all the remaining species by the Ural mountains of central Russia, where it comes partially into contact with other central Siberian taxa (Després, Gielly et al. 2003). Another of its special characteristics is that it displays an extremely distinctive flower morphology, composed of globose and closed yellow flowers, which contrast with the open-flower habit found in other *Trollius* species (Figure 3b, Pellmyr 1989). This latter trait is of particular interest for evolutionary biologists and ecologists who identified it as a character probably resulting of tight coevolutionary processes with associated pollinators (Pellmyr 1992). As a consequence of the closed nature of the flower cup, non-zoophilous (*e.g.,* wind) pollination is impossible

and the flower completely relies on insects for pollen transport, which is especially important considering its self-incompatibility (Pellmyr 1989). Moreover, because of the closed shape of the flower, the number of generalist visitors is strongly reduced, since it is difficult for them to reach the sources of nectar and pollen.

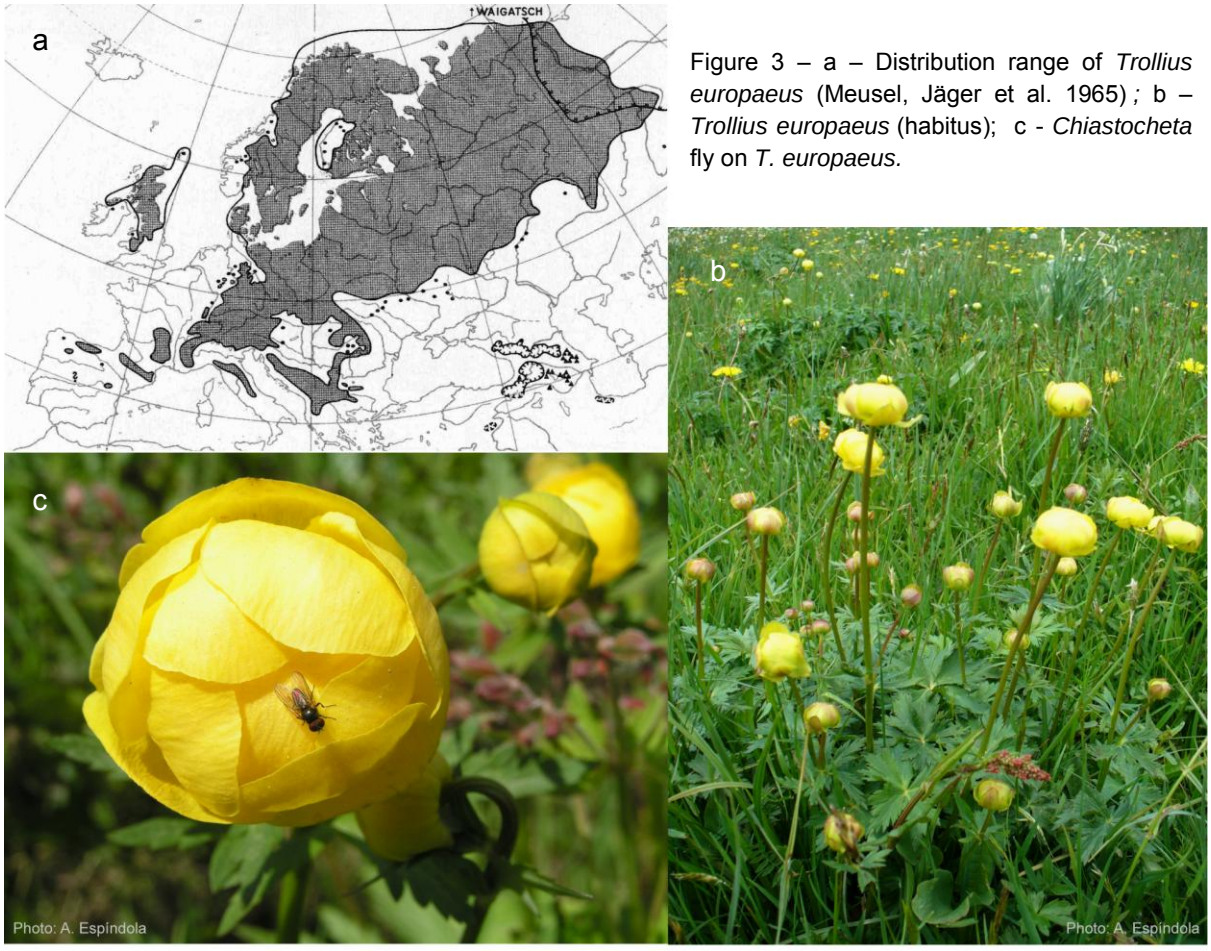


Figure 3 – a – Distribution range of *Trollius europaeus* (Meusel, Jäger et al. 1965) ; b – *Trollius europaeus* (habitus); c - *Chiastocheta* fly on *T. europaeus.*

Starting with Pellmyr (1989), investigations have gathered information that tended to confirm the coevolutionary nature of the forces that shaped this special floral display. For the last three decades, various researchers have demonstrated that the flower shape acts a "filter" for insects not belonging to the genus *Chiastocheta* (Figure 3c, Ibanez, Dujardin et al. 2009). Flies of this genus enter the closed flower to feed on nectar and pollen, mate and passively pollinate female flowers while getting covered with pollen from the enormous anthers. In exchange of the pollination service, flies lay eggs on the carpels into which larvae enter to feed on seeds (Pellmyr 1989). Up to now, eight species (*rotundiventris,*

abruptiventris, dentifera, lophota, inermella, setifera, trollii and *macropyga*) have been demonstrated to exclusively visit *T. europaeus* (Pellmyr 1992; Jaeger and Després 1998).

Research on the biology of these species has also demonstrated that fly species present different and specific behaviors in terms of the position and timing of their laying strategy on the flower carpels. A temporal *suite* of fly visits is observed, with *C. rotundiventris* arriving at the very beginning of the flowering period, and laying generally one egg per flower head (Després and Jaeger 1999). Arriving at the very end of flowering, *C. dentifera* lays several eggs per flower head, ovipositing them between the central carpels (Pellmyr 1992; Després and Jaeger 1999). The remaining species arrive at different intermediate temporal points, laying also several eggs per inflorescence. Besides the fact that adults from different species behave differently during oviposition visits (Després and Jaeger 1999), larvae also differ in their exploitation of the seed resources (Pompanon, Pettex et al. 2006). It has been shown that probably because of an interaction between inter-larvae competition (Després and Jaeger 1999) and larvae-induced toxin accumulation in the plant (Ibanez, Gallet et al. 2009), each species presents a different seed consumption pattern. *C. rotundiventris* for example attacks several carpels and finally develops at the basis of the infructescence, and other species such as *C. dentifera* feeds on all seeds in one carpel (Pellmyr 1989). Once larvae finish feeding, they exit the fruit and fall onto the ground, were they pupate and overwinter (Ibanez, Gallet et al. 2009).

5b. A case study of antagonistic interaction – Arum maculatum L. and Psychodid flies

The most frequent antagonistic plant-insect interactions are composed of phytophagous insects that attack plants. However, rarer cases are those in which plants benefit from the interaction while insects are exploited. Among these rare cases are those involving lure-and-trap pollination. In these interactions, the plant attracts one or several insects to exploit them as pollen carriers. Examples of this type of exploitation are present in several species within

the genera *Aristolochia* L. (Berjano, Ortiz et al. 2009) and *Arum* L. (Gibernau, Macquart et al. 2004), both present in the West-Palearctic.

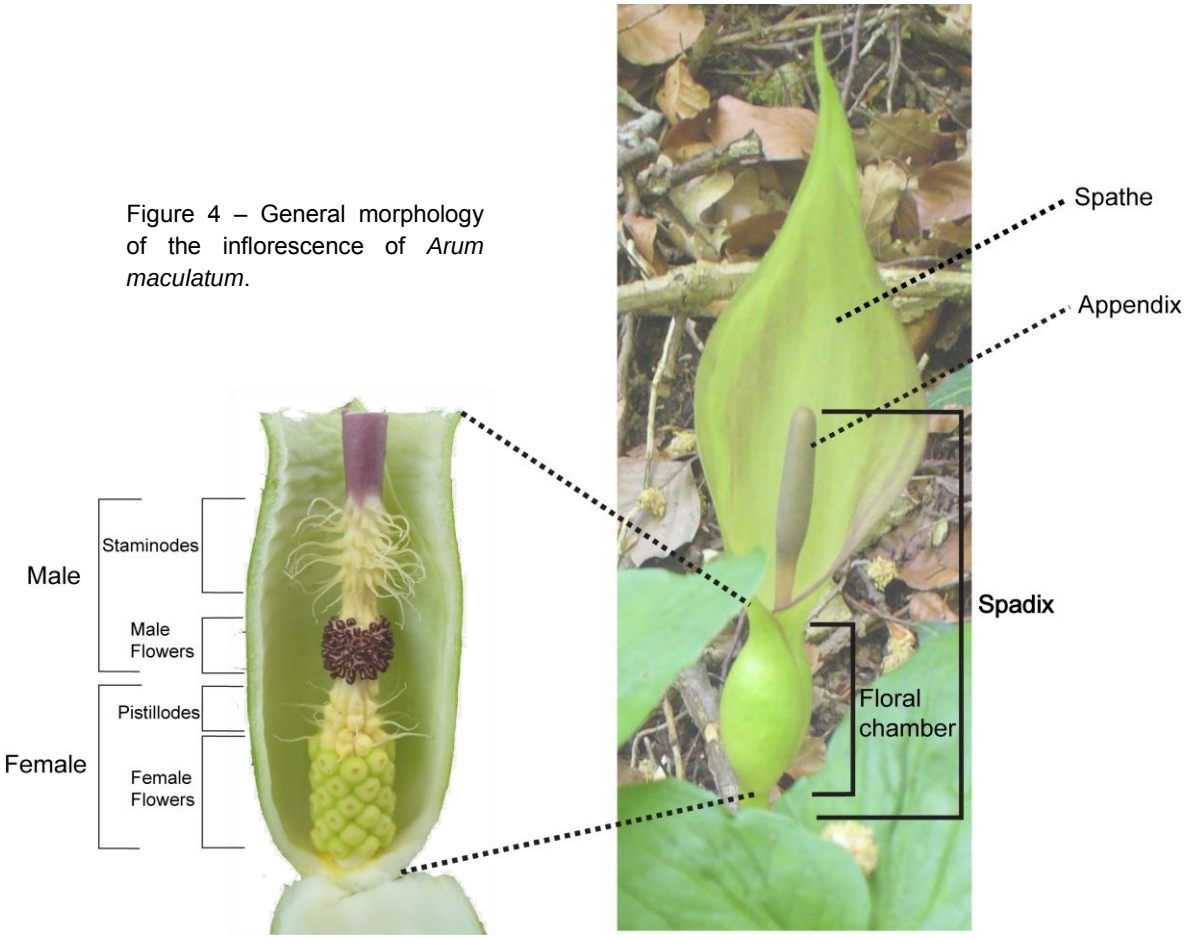


Figure 4 – General morphology of the inflorescence of *Arum maculatum*.

Unlike *Aristolochia*, the genus *Arum* (Araceae) is restricted to a peri-Mediterranean/Eurasian distribution (Boyce 1993; Boyce 2006). This genus is historically emblematic, because of its shape, biology and ecology. Its first mention goes back to ancient Greek records (Theophrastus 370 B.C.), being much later defined by Linnaeus, and currently providing the name of the family. All taxa included in this genus are geophytes and develop underground tubers (Boyce, 1993). As all Araceae, they harbor a special inflorescence morphology (Figure 4), formed by a spathe –a modified bract– that encloses a central spadix, harboring the sexual structures (Mayo, Bogner et al. 1997). In the genus *Arum*, the spadix is separated in two regions: a lower part, which displays female and male flowers, enclosed in the floral chamber –a closed space defined by a contraction of the lower part of the spathe– and an upper part, which is composed of the appendix, responsible of

heat and volatile production (see below) (Figure 4, Boyce 1993). Despite that the genus has been known for centuries, several systematic, historical and geographical issues blurred delimitation and definition of species (Boyce 1993; Bedalov and Küpfer 2005; Boyce 2006).

Among all species within the genus *Arum*, the only one presenting a wide pan-European distribution is *A. maculatum* L. (Figure 5a; *A. italicum* and *A. cylindraceum* are also widely distributed but less northward). Besides being used for centuries as a source of starch, active principles in medicine and for magic grounds (Prime 1960), the species has interested evolutionary biologists because of its very special pollinator interaction (Lack and Diaz 1991). This species displays a specific and obligate lure-and-trap sapromyiophilous pollination system that makes it dependent for cross-pollination on small Psychodid fly species (Figure 5b; Diaz and Kite 2002). Pollination is possible thanks to a series of adaptations displayed by the plant, which allow attracting and trapping ready-to-lay female flies (see below).

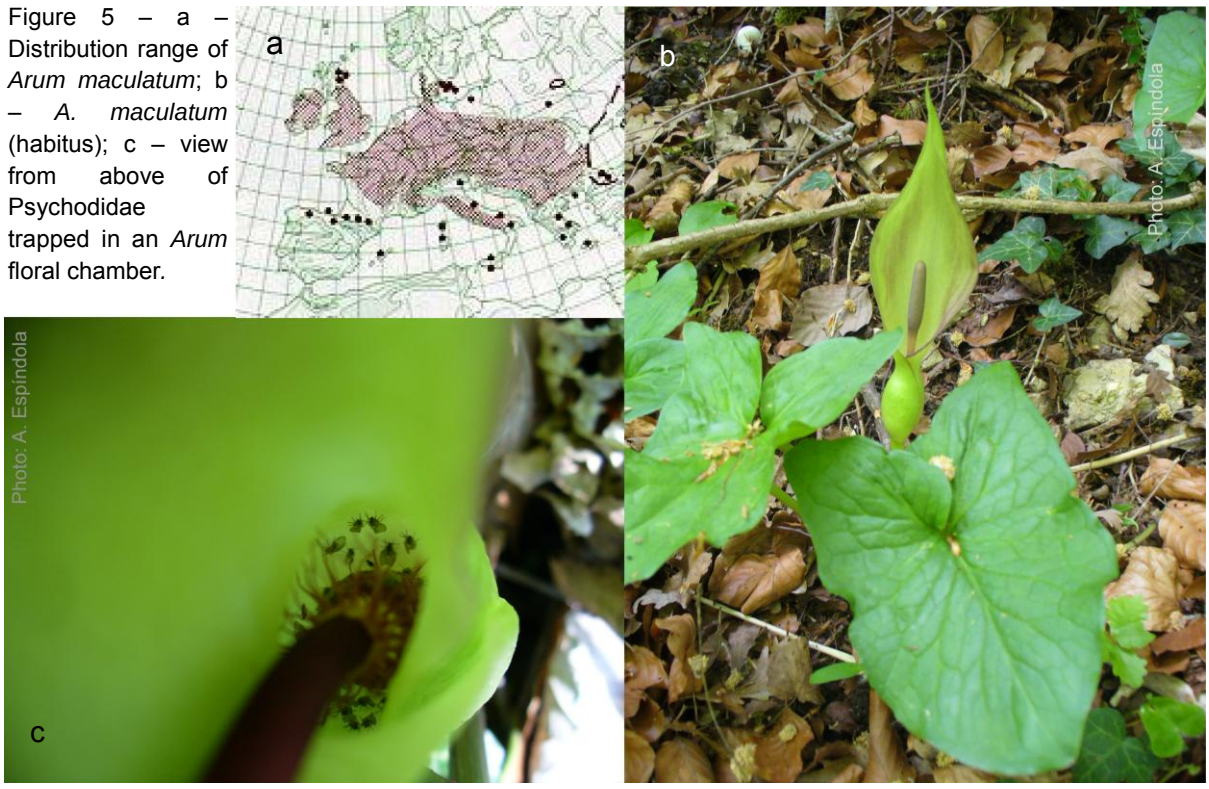


Figure 5 – a – Distribution range of *Arum maculatum*; b – *A. maculatum* (habitus); c – view from above of Psychodidae trapped in an *Arum* floral chamber.

Probably one of the most impressive adaptations is the one involving heat and volatile production. From one day prior to flowering to the second day of anthesis, *A. maculatum*

experiences several cycles of heat production, which can make the flower reach temperatures of up to 15-25°C higher than the surrounding air (Gibernau, Macquart et al. 2004). The first cycle happens the day before anthesis, and is supposed to be involved in female flower maturation. The second helps the spathe unfold and happens at the beginning of flowering. At this moment, volatiles are released from the top of the appendix in association with heat production. Volatile analyses have identified several majoritary compounds being largely present in animal faeces and urine (indole, *p*-cresol, 2-heptanone), and supposed to be key in the attraction of these coprophilous pollinators, whose females are seeking a laying site (Kite 1995).

Flies are thus lured and after landing on the oily spathe, slip into the floral chamber, where they stay trapped because of an array of male flowers modified into hairs (staminodes) that block the exit, and the presence of oil droplets on the chamber walls. Flies are kept prisoners for up to one day, a considerable amount of time if we consider the fly life-span (eight to twelve days; Vaillant 1971). By trying to flee, captured flies will –if they had been previously trapped by another *Arum*– pollinate the female flowers and, once pollen is ripe, they will be covered with it. After liberation of pollen, staminodes wither, walls lose their oily nature and flies are set free. Even if only a very small proportion of them will be trapped again by another *Arum*, this small part will be sufficient to pollinate all flowers and to allow cross-pollination in this species (Lack and Diaz 1991).

One might ask how this type of interaction is maintained through evolutionary times. Several factors could allow keeping up this interaction. On the one hand, fly species involved present an extremely short generation time (12-18 generations/year, Vaillant 1971), which means that only one or two generations are really selected by the plant, which massively flowers during a short period of time (about one month, Ollerton and Diaz 1999). This could indicate that even if selection is strong over one or two fly generations, its effects could be easily lost during the dozen additional generations occurring over the rest of the season. On the other hand, we could also ask why flies continue to be attracted by an odour that they

could avoid; a possible explanation invokes a scenario in which the chemical cue used by the plant to attract the insects is too selective to be counter-selected by them since most odors involved in the floral scents are present in the flies' laying sites. Finally, it is also possible that because insects rarely die in the floral chamber and because population sizes are huge (up to several thousands per square meter, Arshad and Moh Leng 1991), the strength of selection is not substantial compared to the effect of drift.

6. The structure of this study

The present work is divided in two parts, which investigate the different aspects presented above using, as case studies, the two models of specific and obligate plant-insect pollination interactions. The **first section** considers different topics related with *Arum maculatum* L. and its Psychodid pollinators. Here I first investigate, using phylogenetics, biogeography and character tracing, the systematics of the genus *Arum* (**chapter one**). Afterwards, I treat *Arum* phylogeny's counterparts, by identifying the phylogenetic relationships among members of the subfamily Psychodinae (**chapter two**). Further, I proceed to the analysis of the variation in the composition of *A. maculatum'*s visitors at different spatial scales (**chapter three**). Extending the analysis of the plant-insect interaction to a larger view, I evaluate the effects of the identity and abundances of pollinators on the reproductive strategy of two closely related *Arum* species, which strongly differ in their reproductive strategies (**chapter four**). Based on results provided by all previous chapters, I perform phylogeographic analyses of each member of the interaction to finally evaluate the level of congruence among them (**chapter five**).

In the **second section** I study topics related with those of the first section, but on a mutualistic biological model, *i.e.*, the nursery pollination system composed by *T. europaeus* and its *Chiastocheta* pollinators. Because when studying coevolutionary processes it is key to understand the identity and limits of the units under selection, I start this section by evaluating the current species circumscription in the genus *Chiastocheta* (**chapter six**). Then, I evaluate the phylogeography of *T. europaeus,* proposing a new technique to infer

the fate of genetic lineages after distributional contractions in cold-adapted species (**chapter seven)**. Finally, I identify the flies' phylogeographies, which I further compare to the previously obtained spatial genetic structure of the plant (**chapter eight**).

Chapter one – Phylogeny, biogeography and systematics in the genus *Arum* L.

The circumscription of species in the genus *Arum* has been discussed for several years. One of the main difficulties in its systematics lays in the fact that traits in this genus are extremely plastic, which does not help in the identification of morphological synapomorphies. Moreover, because of historical reasons, several taxa have been simultaneously described in different regions under different names, whioch complicated even more the interpretation of the identity of species. In this chapter, using chloroplastic sequences and different analytical approaches, I aim at resolving the infra-generic phylogeny of the genus *Arum*. Considering the need to identify characters informative of the relationships among species, morphological characters are traced on the molecular phylogeny. Finally, I apply biogeographic analyses to identify the most probable center of origin of the genus and to infer the timing and spatialization of dispersal and radiation of the main lineages.

While some subgenera appear supported, most of them are challenged by our results. Phylogenetic inferences thus suggest that taxonomic revisions should be done in this genus. Probably one of the most interesting information provided by this study is related to the identity of *A. maculatum*, which appears separated into two different clades, strongly structured geographically. It thus seems that what was up to now identified as a member of this species to the east of Southern Bulgaria is another specific entity, for which further descriptions should be done in the future. The distribution range of *A. maculatum sensu stricto* is thus strictly European. By tracing morphological characters, we demonstrate that the level of ploidy and the tuber shape are informative of the relationships in the group. It is important to note that this result shows that characters used up to now to define subgeneric

systematic, which are only based on floral measures, are highly homoplasic: because flower characters are highly adapted to the pollination system in *Arum*, it is likely that they evolve extremely quickly with a high rate of homoplasy. Finally, biogeographic analyses demonstrate that the history of the genus is tightly related to the history of the Mediterranean Basin, with colonization and vicariance events happening synchronously with timing of geographical connections and disruptions that occurred in this area.

Chapter two – Phylogenetic relationships within the subfamily Psychodinae in a temporal framework (Diptera: Psychodidae)

Among-family taxonomic relationships in the order Diptera were continuously challenged in the last decades. Such a lack of clear systematics is also found at lower taxonomic levels, as for instance in the subfamily Psychodinae. Molecular phylogeny is a field that allows inferring species relationships without any prior consideration of morphological synapomorphies, which can be further identified once species groups have been settled. Here, I present a molecular phylogeny of this worldwide and very common subfamily, comprising notably species involved in a tight pollination interactions with European Araceae.

Besides showing that the two most abundant visitors of *A. maculatum* correspond to sister lineages distributed in the same sub-clade, we demonstrate that the current relationships in the group need to be strongly reconsidered, arguing for revised and new subdivisions of tribes, subtribes and genera. Moreover, our phylogenetic hypothesis shows that the timing of divergence in Psychodinae spanned the last 86 million years, an ancient temporal origin when investigating divergence time within subfamilies. Similarly, most species originated during the Paleogene (65.5-23 Mya), through a diversification process that followed the Cretaceous-Paleogene (K-T) massive extinction event.

Chapter three – Biogeography of *Arum maculatum*'s visitors

In the last century, *A. maculatum* has been described as a species specifically pollinated by one single fly taxon: *Psychoda phalaenoides* L. (Diptera: Psychodidae). However, studies were performed only in the western part of the plant's distribution range, while *A. maculatum* is widely distributed from southern Scandinavia to Bulgaria and from Spain to Poland. Even though these studies pioneered our knowledge on this antagonistic interaction, they provided only little insights in the consistency of specificity across the distribution range of the plant. By investigating the visitor composition throughout the whole distribution range of the plant, we test predictions from the theory of the Geographic Mosaic of Coevolution (GMC; Thompson 2005), which forecasts that some variation in the nature of interacting species occurs if the environment is heterogeneous.

In this study we sampled floral chambers from all over the distribution range of the plant and identified the variation in flower visitors. In order to quantify the environmental variation related to these sites, we extracted the climatic characteristics of each sampled location from climatic layers and performed ecological niche models analyses. We demonstrate that the composition of the flower chambers varied as a function of latitude, with the most abundant species (*P. phalaenoides*) dominating in central and north-western Europe, while it is almost absent in more Mediterranean environments, where the most abundant species is *Psycha grisescens* Tonnoir, a sister species of *P. phalaenoides*. We show that differences in the composition and relative abundances of one and another species are correlated with environmental factors. Our study thus shows that predictions of the GMC are confirmed in the case of *A. maculatum.*

Chapter four – The effect of insects in the reproductive strategies of three closely related *Arum* species

In coevolutionary interactions, it has been proposed that synergistic forces induce character coevolution in all species involved, which increases the fitness of each organism. The reproductive strategy of a plant can be considered as one of these selected characters. In Araceae, a remarkable reproductive strategy is the one defined by the application of the size-advantage model (SAM). SAM proposes that the energy allocated to different sexes is expected to vary in hermaphrodite organisms with the accumulation of energy in the organism, favoring an investment on the most costly sex when the organism has accumulated enough energy. The application of the SAM has already been investigated in some Araceae, and it has been demonstrated that, while it is applied in some species, this is not necessarily the rule for all species. In the genus *Arum*, more precisely, its application varies depending of the species considered. Former studies showed its application in *A. italicum*, but not in *A. maculatum.*

In this chapter, we propose that the application or not of the SAM could be correlated with the pollination strategy of the plant. In order to test this theory, we compared the pollinator composition of *A. maculatum* (a specialist) with those of *A. cylindraceum* and *A. italicum* (generalists). Based on the idea stated above, we make predictions of application or not of the SAM based on a simple model representing the putative relationship with pollinating insects (specialists vs. generalists) and perform allometric measures in the widespread, but neglected, *A. cylindraceum* to confirm or reject predictions.

Our results indicate that while *A. maculatum* displays pollination features that do not favor the application of the SAM as an adaptive trait, this is probably not the case in other *Arum* presenting less specialized reproductive biologies such as *A. cylindraceum*. Our simple predictive model seems to allow understanding the relationship between pollination strategy and application of the SAM.

Chapter five – Comparative phylogeography of *Arum maculatum* and its specific and obligate pollinators

One of the general objectives of my thesis was to investigate if the phylogeographic patterns of interacting species demonstrated remarkable characteristics. I aimed at testing whether or not the type of interaction could have an effect on the congruence in phylogeographies of associated species, both in mutualistic and in antagonistic specific and obligate interactions.

In this chapter we proceed to the comparison of the phylogeographies of species involved in the antagonistic interaction featuring *A. maculatum* and its pollinating Psychodids. Our working hypothesis is that since in this antagonistic relationship only one of the partners is ecologically dependent of the other, we expect that phylogeographic patterns of the plant and the insects should not be congruent.

In order to test this, we respectively sequenced and applied genomic screenings to insects and plants that we collected at the same locations throughout the whole distribution range of the plant. We afterwards inferred the genetic structure of both actors, identified their spatial distributions, and finally compared the resulting phylogeographic patterns.

Our study reveals that while the plant is clearly geographically structured, this is not the case of the flies. This lack of congruence between the two organisms confirms our initial expectation. It appears that when both species do not reciprocally need each other to survive, phylogeographic patterns are not congruent. Another explanation is however related to the generation times and population sizes present in each species: while the plant has a generation time of 10-15 years, flies undergo 12-18 generations per season, leading to extreme differences in the absolute rate of molecular evolution in both groups. Moreover, insects appear to be able to disperse over large distances, which might blur any phylogeographic signal.

Despite these results, it is important to point out that the genetic structure we obtained for the plant is almost identical to that previously obtained by the analysis of the variation in the composition of the visitors collected in floral chambers over a large spatial scale (chapter three). It is thus possible that some genetic structuring of the plant is driven by the insect species composition, or the other way around.

Chapter six – Phylogeny of *Chiastocheta*

It is not possible to work on coevolutionary systems if the limits of the units of selection are not well defined. In this chapter, we investigated the species concept and limits in *Chiastocheta* species associated with *T. europaeus*. Using different molecular markers and applying biogeographic approaches, we elucidate the phylogenetic relationships among *Chiastocheta* species. The main result of this study is that not all the morphologically defined species are supported by molecular data. While three species show an agreement between morphological and molecular synapomorphies, the others do not present such congruence. We further demonstrate that most species are currently experiencing hybridization, which blurs among-taxon boundaries. The fact that some of these species are hybridizing changes our view of this system, since if hybridization is possible, any coevolutionary force acting on these entities would be much more diffuse than previously thought. From a spatio-temporal point of view, we show that the genus is relatively ancient and that the whole infra-generic diversification occurred during the Quaternary. We also demonstrate that the most recent common ancestor of the genus was actually distributed in Europe, and that dispersal towards Central Asia independently occurred two or three times. Further dispersal towards East Asia and Japan occurred once each from Central Asia.

Chapter seven – Phylogeography of *Trollius europaeus* at the European scale

Little is known about phylogeographic patterns of cold-related organisms in Europe. Moreover, because these organisms present a currently contracted distribution range, the only processes that can be identified when studying their spatial genetic structure are those related to the dynamics of contraction and to range shifts. However, if our aim is to identify such processes, we need to have a view on past genetic diversities to compare past and current patterns. Generally this sight into past diversity is done by using ancient DNA found in fossils or by the inference of past distributions based on macroremains. Unfortunately, ancient material is not systematically available for all species, which restricts our capability to infer the genetic consequences of past demographic contractions related to the termination of glacial periods in cold-related species.

In this chapter we propose a solution to this problem by exploiting the potentials of hindcasting techniques to estimate the past distribution range of *Trollius europaeus* at the LGM. This indirect approach provides a probabilistic view into the past genetic diversity, which is especially useful when considering species for which no ancient material is known. Combining this modeling approach with the analysis of the current genetic structure of the plant, we demonstrate here that the termination of the LGM probably induced a relatively limited loss of genetic diversity in *T. europaeus*, which contrasts with other arctic-alpine species, for which the loss was much more important.

Chapter eight – Phylogeography of three *Chiastocheta* species and the influence of the plant in shaping their genetic structure

The aims of this chapter are to test if closely ecologically related species present similar phylogeographic patterns, and to understand what is the level of congruence between the phylogeographic patterns of the pollinators and the plant in the *Trollius*-*Chiastocheta* mutualistic interaction.

Based on results of chapter six, we investigate the phylogeographic patterns of the three species that appeared to be well-defined both morphologically and genetically in our previous molecular analysis (*i.e.*, *C. rotundiventris, C. lophota* and *C. dentifera*). To do so, we used sequence data to infer phylogenetic topologies. On the one hand we applied "traditional" phylogeographic methods, in which we visually estimated the geographic distribution of the genetic clades and clusters in space. Second, we applied statistical phylogeography techniques, which allowed us to directly test the probability that the genetic structure obtained in the phylogenetic inferences has been produced under a given coalescent scenario defined by the plant genetic structure (chapter seven).

Our results indicate that, despite having similar ecologies, each fly species presents a different phylogeographic pattern, with one species showing a clearly geographically defined genetic structure (*C. lophota*), another harboring very high levels of admixture (*C. rotundiventris*), and the third presenting a phylogeographic structure which does not statistically differ of what would have been obtained in the case that the plant's spatial genetic structure had driven its phylogeography (*C. dentifera*). We thus demonstrate that the phylogeographic histories of at least one fly species and that of the plant statistically overlap, which partially confirms our initial working hypothesis of congruence in phylogeographic patterns of species involved in mutualistic interactions.

7. References

Aizen, M. A. and C. Ezcurra (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." Ecología Austral **8**: 217-236.

Aizen, M. A. and P. Feinsinger (1994). "Forest fragmentation, pollination, and plant reproduction in a Chaco dry forest, Argentina." Ecology **75**: 330-351.

Alvarez, N., C. Thiel-Egenter, et al. (2009). "History or ecology? Substrate type as a major driver of spatial genetic structure in Alpine plants." Ecology Letters **12**(7): 632-40.

Arshad, A. and K.-Y. Moh Leng (1991). "Preliminary Population Assessment of *Psychoda alternata* (Diptera: Psychodidae) in SoilIrrigated with Wastewater for Turf Cultivation." The Florida Entomologist **74**(4): 591-596.

Avise, J. C. (2009). "Phylogeography: retrospect and prospect." Journal of Biogeography **36**: 3-15.

Bedalov, M. and P. Küpfer (2005). "Studies on the genus *Arum* (Araceae)." Bulletin de la société Neuchâteloise des Sciences Naturelles **128**: 43-70.

Berjano, R., P. L. Ortiz, et al. (2009). "Pollinators, flowering phenology and floral longevity in two Mediterranean Aristolochia species, with a review of flower visitor records for the genus." Plant Biology **11**(1): 6-16.

Boyce, P. C. (1993). The genus *Arum*. London, UK, The Royal Botanical Gardens, Kew.

Boyce, P. C. (2006). "*Arum* - a decade of change." Aroideana **29**: 132-137.

Brochmann, C., T. M. Gabrielsen, et al. (2003). "Glacial survival or *tabula rasa*? The history of North Atlantic biota revisited." Taxon **52**: 417-450.

Bronstein, J. L., U. Dieckmann, et al. (2004). Coevolutionary dynamics and the conservation of mutualisms. I. I. f. A. S. Analysis. Laxenburg, International Institute for Applied Systems Analysis**:** 1-23.

Cheddadi, R., G. G. Vendramin, et al. (2006). "Imprints of glacial refugia in the modern genetic diversity of *Pinus sylvestris*." Global Ecology and Biogeography **15**: 271-282.

Crepet, W. L. (1996). "Timing in the evolution of derived floral characters: upper Cretaceous (Turonian) taxa with tricolpate and tricolpate-derived pollen." Review of Palaeobotany and Palynology.

Crepet, W. L. (2000). "Progress in understanding angiosperm history, success, and relationships: Darwin's abominably "perplexing phenomenon"." Proceedings of the National Academy of Sciences of the United States of America **97**(24): 12939-12941.

Després, L., L. Gielly, et al. (2003). "Using AFLP to resolve phylogenetic relationships in a morphologically diversified plant species complex when nuclear and chloroplast sequences fail to reveal variability." Molecular Phylogenetics and Evolution **27**(2): 185-196.

Després, L. and N. Jaeger (1999). "Evolution of oviposition strategies and speciation in the globeflower flies *Chiastocheta* spp. (Anthomyiidae)." Journal of Evolutionary Biology **12**: 822-831.

Després, L., E. Pettex, et al. (2002). "Speciation in the globeflower fly *Chiastocheta* spp. (Diptera: Anthomyiidae) in relation to host plant species, biogeography, and morphology." Molecular Phylogenetics and Evolution **22**(2): 258-68.

Diaz, A. and G. C. Kite (2002). "A comparison of the pollination ecology of *Arum maculatum* and *A. italicum* in England." Watsonia **24**: 171-181.

Doroszewska, A.-. (1974). "The genus Trollius L., a taxonomical study." Monographie Botanicae **41**: 1-167.

Dufaÿ, M. and M.-C. Anstett (2003). "Conflicts between plants and pollinators that reproduce within inflorescences: evolutionary variations on a theme." OIKOS **100**: 3-14.

Ehlers, J. and P. L. Gibbard (2004). Quaternary glaciations: extent and chronology. Part I: Europe. Amsterdam ; San Diego, Elsevier.

Ehrich, D., M. Gaudeul, et al. (2007). "Genetic consequences of Pleistocene range shifts: contrast between the Arctic, the Alps and the East African mountains." Molecular Ecology **16**(12): 2542-59.

Ehrlich, P. R. and P. H. Raven (1964). "Butterflies and Plants: A study in coevolution." Evolution **18**(4): 586-607.

Farrell, B. D., C. Mitter, et al. (1992). "Diversification at the insect-plant interface. Insights from phylogenetics." BioScience **42**(1): 34-42.

Futuyma, D. J. and G. Moreno (1988). "The evolution of ecological specialization." Annual Review of Ecology and Systematics **19**: 207-233.

Gibernau, M., D. Macquart, et al. (2004). "Pollination in the genus *Arum* - a review." Aroideana **27**: 148-166.

Grimaldi, D. A. (1999). "The co-radiations of pollinating insects and angiosperms in the Cretaceous." Annals of the Missouri Botanical Gardens **86**: 373-406.

Hennig, W., Ed. (1976). Anthomyiidae. Die Fliegen der Palaearktischen Region. Stuttgart.

Hewitt, G. (2000). "The genetic legacy of the Quaternary ice ages." Nature **405**(6789): 907-913.

Hewitt, G. M. (1996). "Some genetic consequences of ice ages, and their role in divergence and speciation." Biological Journal of the Linnean Society **58**(3): 247-276.

Hewitt, G. M. (1999). "Post-glacial re-colonization of European biota." Biological Journal of the Linnean Society **68**(1-2): 87-112.

Hewitt, G. M. (2001). "Speciation, hybrid zones and phylogeography - or seeing genes in space and time." Molecular Ecology **10**(3): 537-549.

Hutchinson, G. E. (1957). Cold Spring Harbor Symposia on Quantitative Biology. Concluding remarks. Cold Spring Harbor Symposia on Quantitative Biology.

Ibanez, S., G. Dujardin, et al. (2009). "Stability of floral specialization in *Trollius europaeus* in contrasting ecological environments." Journal of Evolutionary Biology **22**(6): 1183-92.

Ibanez, S., C. Gallet, et al. (2009). "Plant chemical defence: a partner control mechanism stabilising plant--seed-eating pollinator mutualisms." BMC Evolutionary Biology **9**: 261.

Jaeger, N. and L. Després (1998). "Obligate mutualism between *Trollius europaeus* and its seed-parasite pollinators *Chiastocheta* flies in the Alps." Comptes Rendus De

L'Académie Des Sciences Serie III-Sciences De La Vie-Life Sciences **321**(9): 789-796.

Jousselin, E., S. van Noort, et al. (2008). "One fig to bind them all: host conservatism in a fig wasp community unraveled by cospeciation analyses among pollinating and nonpollinating fig wasps." Evolution **62**(7): 1777-97.

Kephart, S., R. J. Reynolds, et al. (2006). "Pollination and seed predation by moths on *Silene* and allied Caryophyllaceae: evaluating a model system to study the evolution of mutualisms." New Phytologist **169**(4): 667-680.

Kite, G. C. (1995). "The floral odour of *Arum maculatum*." Biochemical Systematics and Ecology **23**: 343-354.

Lack, A. J. and A. Diaz (1991). "The pollination of *Arum maculatum* L. - a historical review and new observations." Watsonia **18**: 333-342.

Lunau, K. (2004). "Adaptive radiation and coevolution - pollination biology case studies." Organisms Diversity & Evolution **4**(3): 207-224.

Margulis, L. (1970). Origin of Eukaryotic Cells. New Haven, Connecticut, USA, Yale University Press.

Mayo, S. J., J. Bogner, et al. (1997). The genera of Araceae. London, UK., Royal Botanical Gardens, Kew.

Meusel, H., E. Jäger, et al. (1965). Vergleichende Chorologie der zentraleuropäischen Flora. Bd.1 (Text und Karten). Jena, Fischer Verlag.

Michelsen, V. (1985). "A revision of the Anthomyiidae (Diptera) described by J. W. Zetterstedt." Steenstrupia **11**(2): 37-65.

Mitter, C., B. D. Farrell, et al. (1991). "Phylogenetic studies of insect-plant interactions: insights into the genesis of diversity." Trends in Ecology & Evolution **6**(9): 290-293.

Ollerton, J. and A. Diaz (1999). "Evidence for stabilising selection acting on flowering time in *Arum maculatum* (Araceae): the influence of phylogeny on adaptation." Oecologia **119**(3): 340-348.

Ollerton, J., A. Killick, et al. (2007). "Multiple meanings and modes: on the many ways to be a generalist flower." Taxon **56**(3): 1-12.

Pellissier, L., N. Alvarez, et al. (in press). Pollinators as drivers of plant distribution and assemblage into communities. Evolution of the Plant Pollinator Relationships. S. Patiny. UK, Cambridge University Press.

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." Biological Journal of the Linnean Society **47**: 337-365.

Pellmyr, O. (2002). Pollination by animals. Plant-animal interactions: an evolutionary approach. C. M. Herrera and O. Pellmyr. Oxford, UK, Blackwell.

Pellmyr, O. and C. J. Huth (1994). "Evolutionary stability of mutualism between yuccas and yucca moths." Nature **372**: 257-260.

Pompanon, F., E. Pettex, et al. (2006). "Patterns of resource exploitation in four coexisting globeflower fly species (*Chiastocheta* sp.)." Acta Oecologica-International Journal of Ecology **29**(2): 233-240.

Prime, C. T. (1960). Lords & Ladies. London, UK, Collins.

Provan, J. and K. D. Bennett (2008). "Phylogeographic insights into cryptic glacial refugia." Trends in Ecology & Evolution **23**(10): 564-71.

Schmitt, T. (2007). "Molecular biogeography of Europe: Pleistocene cycles and postglacial trends." Frontiers in Zoology **4**: 11.

Schönswetter, P., I. Stehlik, et al. (2005). "Molecular evidence for glacial refugia of mountain plants in the European Alps." Molecular Ecology **14**(11): 3547-55.

Schönswetter, P., A. Tribsch, et al. (2004). "Amplified fragment length polymorphism (AFLP) suggests old and recent immigration into the Alps by the arctic-alpine annual *Comastoma tenellum* (Gentianaceae)." Journal of Biogeography **31**: 1673-1681.

Shanahan, M., S. So, et al. (2001). "Fig-eating by vertebrate frugivores: a global review." Biological Reviews of the Cambridge Philosophical Society (London) **76**(4): 529-72.

Smith, T. B., L. A. Freed, et al. (1995). "Evolutionary consequences of extinctions in populations of a Hawaiian honeycreeper." Conservation Biology **9**: 107-113.

Stewart, J. R. and A. M. Lister (2001). "Cryptic northern refugia and the origins of the modern biota." Trends in Ecology & Evolution **16**(11): 608-613.

Stewart, J. R., A. M. Lister, et al. (2010). "Refugia revisited: individualistic responses of species in space and time." Proceedings of the Royal Society of London B **277**: 661-671.

Taberlet, P. and R. Cheddadi (2002). "Quaternary refugia and persistence of biodiversity." Science **297**(5589): 2009-2010.

Taberlet, P., L. Fumagalli, et al. (1998). "Comparative phylogeography and postglacial colonization routes in Europe." Molecular Ecology **7**(4): 453-464.

Theophrastus (370 B.C.). The enquiry into plants. W. Heinemann. London, UK.

Thompson, J. N. (1994). The coevolutionary process. Chicago, University of Chicago Press.

Thompson, J. N. (2005). "Coevolution: the geographic mosaic of coevolutionary arms races." Current Biology **15**(24): 992-994.

Thompson, J. N. (2009). "The coevolving web of life." American Naturalist **173**(2): 125-40.

Tripp, E. A. and P. S. Manos (2008). "Is floral specialization an evolutionary dead-end? Pollination system transitions in Ruellia (Acanthaceae)." Evolution **62**(7): 1712-37.

Tsai, Y.-H. E. and P. S. Manos (2010). "Host density drives the postglacial migration of the tree parasite, *Epifagus virginiana*." Proceedings of the National Academy of Sciences of the United States of America **107**(39): 17035-17040.

Vaillant, F. (1971). 9d. Psychodidae-Psychodinae. Die Fliegen der Palaearktischen Region. E. Lindner. Stuttgart, E. Schweizerbartische Verlagsbuchhandlung. **9d:** 358.

Velichko, A. A., E. Y. Novenko, et al. (2007). Comparative Analysis of Vegetation and Climate Changes During the Eemian Interglacial in Central and Eastern Europe. The climate of past interglacials. F. Sirocko, M. Claussen, M. F. Sánchez Goñi and T. Litt. Amsterdam, The Netherlands, Elsevier**:** 255-264.

Walther, G. R., E. Post, et al. (2002). "Ecological responses to recent climate change." Nature **416**(6879): 389-95.

Chapter one

New insights into the phylogenetics and biogeography of Arum (Araceae): unravelling its evolutionary history

Anahí Espíndola, Sven Buerki, Marija Bedalov,
Philippe Küpfer and Nadir Alvarez

Published in the *Botanical Journal of the Linnean Society*

Published in Botanical Journal of the Linnean Society 163, issue 1, 14-32, 2010
which should be used for any reference to this work

New insights into the phylogenetics and biogeography of *Arum* (Araceae): unravelling its evolutionary history

ANAHÍ ESPÍNDOLA[1]*, SVEN BUERKI[2], MARIJA BEDALOV[3,4], PHILIPPE KÜPFER[4] and NADIR ALVAREZ[5]

[1]*Laboratory of Evolutionary Entomology, Institute of Biology, University of Neuchâtel, Rue Emile-Argand 11, CH-2009 Neuchâtel, Switzerland*
[2]*Department of Biodiversity and Conservation, Real Jardín Botánico, CSIC, Plaza de Murillo 2, 28014 Madrid, Spain*
[3]*Department of Botany, Faculty of Sciences, University of Zagreb, Marulićev Trg 20/II, HR-10000 Zagreb, Croatia*
[4]*Laboratory of Evolutionary Botany, Institute of Biology, University of Neuchâtel, Rue Emile-Argand 11, CH-2009 Neuchâtel, Switzerland*
[5]*Department of Ecology and Evolution, University of Lausanne, Biophore Building, 1015 Lausanne, Switzerland*

The heat- and odour-producing genus *Arum* (Araceae) has interested scientists for centuries. This long-term interest has allowed a deep knowledge of some complex processes, such as the physiology and dynamics of its characteristic lure-and-trap pollination system, to be built up. However, mainly because of its large distributional range and high degree of morphological variation, species' limits and relationships are still under discussion. Today, the genus comprises 28 species subdivided into two subgenera, two sections and six subsections. In this study, the phylogeny of the genus is inferred on the basis of four plastid regions, and the evolution of several morphological characters is investigated. Our phylogenetic hypothesis is not in agreement with the current infrageneric classification of the genus and challenges the monophyly of several species. This demonstrates the need for a new infrageneric classification based on characters reflecting the evolution of this enigmatic genus. To investigate the biogeography of *Arum* deeply, further spatiotemporal analyses were performed, addressing the importance of the Mediterranean basin in the diversification of *Arum*. Our results suggest that its centre of origin was the European–Aegean region, and that major diversification happened during the last 10 Myr.

ADDITIONAL KEYWORDS: character tracing – infrageneric systematics – Mediterranean biogeography – phylogenetic inferences.

INTRODUCTION

With 109 genera and over 3700 species described (Mayo, Bogner & Boyce, 1997) Araceae have a worldwide distribution and are found in a wide range of environments, from Arctic–Alpine (e.g. *Calla palustris* L.) to xerophytic (e.g. *Anthurium nizandense* Matuda), with most species occurring in the tropics. The family encompasses a large variety of life forms, from epiphytic to aquatic, attesting the numerous adaptive radiations that have occurred in this early Cretaceous family (Chase *et al*., 2006; Anderson & Janssen, 2009). A remarkable feature in Araceae is the evolution of heat production in several genera (Minorsky, 2003), especially those displaying pollination-related associations with arthropods, in which thermogenesis is associated with the emission of volatile compounds and the attraction of pollinators (Moodie, 1975).

*Corresponding author. E-mail: maria.espindola@unine.ch

One of the few Palaearctic representatives of Araceae is the herbaceous genus *Arum* L., which comprises 28 described species (Lobin *et al*., 2007; CATE project, 2010). Because of its characteristic flowers, showing adaptations for trapping pollinators (Gibernau, Macquart & Przetak, 2004), and its ability to produce heat and odours, this genus has fascinated not only scientists but also the wider public for centuries (Theophrastus, 370 BC; Hruby, 1910; Boyce, 1993). As shown by archaeological and historical records, several *Arum* spp. have been used by humans since ancient times for food (tubers), medicine (fruits, leaves, tubers), fashion (tuber starch) and even magic (leaf decoctions) (Prime, 1960).

The inflorescences of *Arum* consist of two parts: a spadix and a spathe (Boyce, 1993). The spadix displays the unisexual flowers and harbours adaptations involved in heat production, whereas the spathe is a modified bract surrounding the spadix. One of the distinctive synapomorphies of the genus is the partition of the spadix. The lower zone comprises both female (lower portion) and male (upper portion) flowers placed in a floral chamber, which is usually delimited by male-sterile flowers modified as hairs: the staminodes. Its apex is a smooth, subcylindrical, usually stipitate organ, known as the spadix appendix (Boyce, 1993). This structure is also recognized as an efficient thermogenetic organ with which the plant attracts pollinating arthropods with heat and production of volatile compounds. The combination of odour emission and hair presence at the top of the floral chamber (acting as a fence) is a key feature for the efficient trapping of arthropods during the female receptive period and until pollen release (Gibernau *et al*., 2004).

Historically, the genus was defined by Fuchs (1542) and later established by Linnaeus (1753). The different species were first circumscribed on the basis of morphology (Schott, 1832), and chromosome counts (Bedalov, 1981) led to the identification of different ploidies in the genus (di-, tetra- and hexaploids, $x = 14$; for a review, see Boyce, 1989). In the most recent revisions of the genus (Boyce, 1993, 1994, 2006; Bedalov & Küpfer, 2005), several morphological characters (tuber shape, flower disposition, growth period, spadix shape and structure of sterile flowers) have been used to build a classification comprising two subgenera, two sections and six subsections. The subgenus *Gymnomesium* (Schott) Engl. is monospecific, including only the Hercynian endemic *Arum pictum* L.f. Subgenus *Arum* Engl. includes sections *Arum* and *Dioscoridea* Bronner, the latter being divided into six subsections (Table 1).

This classification may be controversial, notably because: (1) several taxa have been defined on the basis of herbarium specimens (this approach may not be optimal in this group as important characters are observable only on fresh material; Boyce, 1989); (2) species having large distributions and studied locally were sometimes either simultaneously described under different names (e.g. *A. italicum* Mill.) or assigned to different taxa when they belonged to the same taxon (e.g. *A. cylindraceum* Gasp.) (Bedalov & Küpfer, 2005); (3) following this last point, as several species harbour a high level of intraspecific polymorphism, this may even trigger the splitting of widely distributed taxa (Boyce, 2006). Therefore, it is now an appropriate time to evaluate the systematics of *Arum* based on molecular evidence. Published molecular phylogenetic analyses including species of *Arum* have focused on the investigation of relationships at the family level and have lacked sampling and resolution at the infrageneric level (Cabrera *et al*., 2008; Mansion *et al*., 2008). In this article, we aim to produce a phylogenetic hypothesis for the genus *Arum* by sequencing four plastid regions suitable for addressing relationships at the infrageneric level based on 26 of the 28 described species. This will allow us to assess the validity of the current classification and to examine the evolution of several key characters. To decipher the evolutionary history of this early Miocene genus (Mansion *et al*., 2008), we perform spatiotemporal analyses to determine the events that played a central role in the radiation, dispersion and isolation of the different species (Sanmartín, Enghoff & Ronquist, 2001; Médail & Diadema, 2009). Finally, on the basis of our results, we suggest guidelines for a new infrageneric classification of the genus *Arum*.

MATERIAL AND METHODS

SAMPLING

Analyses were based on 64 specimens, representing 26 of the 28 described species and spanning all subgenera, sections and subsections of *Arum* (Table 1). On the basis of Mansion *et al*. (2008), *Dracunculus canariensis* Kunth, *D. vulgaris* Schott, *Biarum davisii* Turrill and *B. dispar* (Schott) Talavera were used as outgroup taxa. Samples were either provided by the DNA Bank of the Royal Botanic Gardens, Kew (UK) or extracted directly from dried plant material from herbaria or field collections (Appendix 1).

DNA EXTRACTION, AMPLIFICATION AND SEQUENCING

DNA of freshly collected material and herbarium samples was extracted using the DNeasy Plant Kit (Qiagen, Basle, Switzerland). The plastid regions *3′rps16-5′trnK*, *ndhA* intron, *psbD-trnT* and *rpl32-trnL* were amplified with the primers described in Shaw *et al*. (2007). Amplifications were performed in a

Table 1. Current subgeneric taxonomy of genus *Arum* L. Taxa with an '*' were not included in the present study.

Subgenus	Section	Subsection	Species
Gymnomesium			*A. pictum* L.f.
Arum	*Arum*		*A. byzantinum* Schott
			A. concinnatum Schott
			A. italicum Mill.
			A. maculatum L.
			A. megobrebi Lobin, M.Neumann, Bogner & P.C.Boyce
	Dioscoridea	*Alpina*	*A. cylindraceum* Gasp.
			A. lucanum Cavara & Grande
		Discroochiton	*A. apulum* (Carano) P.Boyce
			A. balansanum R.R.Mill
			A. besserianum Schott
			A. cyrenaicum Hruby
			A. elongatum Steven
			A. gratum Schott*
			A. hainesii Riedl*
			A. nigrum Vell.
			A. orientale M.Bieb.
			A. purpureospathum P.C.Boyce
			A. sintenisii P.C.Boyce
		Tenuifila	*A. jacquemontii* Blume
			A. korolkowii Regel
			A. rupicola Boiss.
		Hygrophila	*A. euxinum* R.R.Mill
			A. hygrophilum Boiss.
		Poeciloporphyrochiton	*A. dioscoridis* Sibth. & Sm.
			A. palaestinum Boiss.
		Cretica	*A. creticum* Boiss. & Heldr.
			A. idaeum Coust. & Gand.

master mix containing 0.5 × buffer, 150 mM deoxynucleoside triphosphate (dNTP), 0.7 mM $MgCl_2$, 0.3 μg μL^{-1} bovine serum albumin (BSA), 0.5 μM primers and 1 unit of *Taq* Polymerase (Promega, Dübendorf, Switzerland) made up to a final volume of 30 μL with purified MilliQ water. Reactions were run in a TGradient thermocycler (Biometra, Goettingen, Germany). Initial denaturation was programmed for 2 min 30 s at 95 °C, followed by 35 cycles at 95 °C for 35 s, 54–60 °C for 45 s, 72 °C for 1 min and a final extension of 8 min at 72 °C. The purification of PCR products and fluorescence sequencing were performed by Macrogen, Inc. (Seoul, South Korea) and Fasteris Life Sciences (Geneva, Switzerland) with the same primers as used for PCR amplification.

Sequence alignment and phylogenetic reconstructions

Automatically generated base-calls for all sequences were checked and edited using ChromasPro 1.41 (Technelysium Pty Ltd, Tewantin, Australia). For each plastid region, alignment was performed using the ClustalW algorithm implemented in Bioedit 7.0 (Hall, 1999), followed by minor manual corrections. After concatenation of the four regions, a matrix of 3723 bp was obtained. Gaps were further coded following the simple method of Simmons & Ochoterena (2000), as implemented in FastGap 1.2 (Borchsenius, 2009).

The numbers of constant (C), variable (V) and potentially parsimony-informative (PI) sites were calculated for each partition using PAUP* v4.0b10 (Swofford, 2002). Before computing total evidence trees, we tested for incongruence among the four regions by applying the partition homogeneity test as implemented in PAUP* v4.0b10 with 100 replicates (this test is equivalent to the incongruence length difference test of Farris *et al*., 1994; for convenience, it is referred to as the ILD test). Total evidence trees (*sensu* Kluge, 1989) were determined using both Bayesian inference and maximum parsimony (MP) approaches.

Bayesian Markov chain Monte Carlo (MCMC) methods were used to approximate the posterior prob-

ability distribution of the phylogenetic trees on the basis of the combined plastid dataset with four distinct partitions plus one partition for the gap information, by running MrBayes v.3.1.2 (Ronquist & Huelsenbeck, 2003). Model selection for the plastid DNA partitions was tested using MrAIC (Nylander, 2004) based on the Akaike information criterion (Akaike, 1973), and a restriction model was applied to the partition containing the coded gaps. Three independent runs with one cold and five heated chains were run for 5×10^7 generations each. Frequencies were sampled every 1000 generations and temperature was fixed to 0.5. The convergence of MCMC was tested by computing the potential scale reduction factor (PSRF; Gelman & Rubin, 1992) as implemented in MrBayes, and by determining the effective sample size for each parameter using Tracer v.1.4 (Rambaut & Drummond, 2004). Accordingly, the burn-in period was set to 10^7 generations until stationarity of the likelihood value was established among the runs, and 10 000 sample points were discarded (20% of the total number of trees). The remaining 40 001 trees from each run were pooled (120 003 trees in total) to estimate the posterior probability distribution of the phylogenetic inference. To yield a single phylogenetic hypothesis, the posterior distribution was summarized in the 50% majority-rule consensus tree (referred to as the half-compatible tree in MrBayes), with a Bayesian posterior probability (BPP) at each node indicating statistical support.

The combined dataset was further analysed under the MP criterion using the parsimony ratchet (Nixon, 1999) as implemented in PAUPrat (Sikes & Lewis, 2001). Based on recommendations by Nixon (1999), ten independent searches were performed with 200 iterations, and 15% of the parsimony-informative characters were perturbed using PAUP* version 4.0b10. The shortest equally most parsimonious trees were combined to produce a strict consensus tree. Node support was determined by computing decay indices (DIs) (Bremer, 1988) as implemented in TreeRot 3.0 (Sorenson & Franzosa, 2007). DI measures the number of extra steps in tree length required before a node collapses (Bremer, 1988; Baker & DeSalle, 1997).

Finally, the level of congruence between Bayesian and MP analyses was determined by computing the quartet distance (Estabrook, 1992) between the two topologies. Considering that the distances between the different topologies were small (see Phylogenetic inferences section in Results), the remaining analyses were only based on the Bayesian inference analysis.

Character evolution

Character tracing was performed on traits generally used in taxonomic studies of *Arum*. On the basis of the topology of the 50% majority-rule Bayesian analysis, the following categorical characters were mapped using Mesquite 2.6 (Maddison & Maddison, 2009) with accelerated transformation optimization (ACCTRAN) and unordered parsimony: tuber form (rhizomatous/discoid), flower type (flag/cryptic), spadix/spathe ratio (0–0.5; 0.5–1; > 1) and ploidy (di-, tetra-, hexaploid). Characters were obtained from the latest systematic studies performed on the species (Boyce, 1993, 2006; Bedalov & Küpfer, 2005; Lobin *et al*., 2007).

Dating and biogeographical analyses

Because the molecular clock hypothesis was rejected (data not shown), the 50% majority-rule Bayesian inference tree was rendered ultrametric using the penalized-likelihood method (Sanderson, 2002; hereafter PL), as implemented in the program r8s v.1.71 (Sanderson, 2004) by applying a smoothing value of 1000 and the truncated Newton algorithm. The most external outgroup, *B. davisii*, was pruned for the estimation of the divergence time as required by the program (see Sanderson, 2004). The following calibration points were applied (according to Mansion *et al*., 2008): (1) the root node (i.e. the most recent common ancestor of genera *Arum*, *Biarum* Schott and *Dracunculus* Mill.) was constrained to a maximum age of 30.2 Mya; (2) the most recent common ancestor of *Arum* and *Dracunculus* was constrained to a minimum age of 27.3 Mya; and (3) the stem group of *Arum* subgenus *Arum* was constrained to a minimum age of 16.1 Mya.

Areas were defined following different studies on the geological and biogeographical history of the Mediterranean basin and surrounding areas (Meulenkamp & Sissingh, 2003; Mansion *et al*., 2008; Ree & Sanmartín, 2009), and were set to a number of ten: East European, West European, Apennines, Aegean, Anatolian, Iranian, Arabian, North African, Macaronesian Islands and Caucasus (Fig. 3). The rules applied to define the area for each species were as follows: (1) if the origin of the sample was known, the sample was attributed to the area in which it was sampled; (2) if the origin was unknown, the sample was assigned to the area in which the plant is known to be distributed according to Boyce (1993, 2006) and the search engine of *Flora Europaea* (Flora Europaea, 2009) [in the case of *A. balansanum* R.R.Mill., *A. byzantinum* Schott, *A. sintenisii* (Engl.) P.C.Boyce, *D. vulgaris*, *D. canariensis* and *B. dispar*]; (3) if a sample belonged to a widely distributed and well-described species for which we did not possess samples from all the parts of the distributional area, it was assigned to its region of origin plus the remaining noncovered regions according to Boyce (1993) (only in the case of *A. italicum*).

Table 2. Sequenced regions, with corresponding total number and percentages of base pairs (bp), constant (C), variable (V) and parsimony informative (PI) sites.

Region	Total (bp)	Constant (C) sites	Variable (V) sites	Parsimony-informative (PI) sites
3′rps16-5′trnK	845 (100%)	791 (93.6%)	54 (6.4%)	17 (2.0%)
ndhA intron	1077 (100%)	1021 (94.8%)	56 (5.2%)	24 (2.2%)
psbD-trnT	1024 (100%)	993 (97.0%)	31 (3.0%)	13 (1.3%)
rpl32-trnL	777 (100%)	730 (94.0%)	47 (6.0%)	36 (4.6%)

Dispersal-vicariance analysis (DIVA) is a method for inferring the most parsimonious reconstruction of ancestral ranges on a given phylogenetic tree by minimizing the number of dispersal and extinction events that are needed to explain the current terminal distributions (Ronquist, 1997). The program DIVA (Ronquist, 2001) uses a three-dimensional cost matrix to estimate the cost of moving from the ancestor to each of the descendants (Ronquist, 1997). It allows two different scenarios for range inheritance at speciation nodes: (1) duplication or within-area speciation, when the ancestor is distributed in a single area and each of the two descendants inherits the entire ancestral range (e.g. A to A); (2) vicariance, when the ancestor occurs in two or more areas and each descendant inherits a nonoverlapping subset of the ancestral range (e.g. AB to A and B). Only one dispersal event per branch (between two ancestral nodes) is allowed in the model, except for terminal branches leading to widespread taxa, for which DIVA postulates multiple dispersal events. To account for polytomies in the 50% majority-rule Bayesian inference tree, five more exceptions were required in our analysis (see below). DIVAs were run with the maximum number of areas allowed at ancestral nodes constrained to two. Uncertainty in phylogenetic relationships was accounted for in DIVA by using an approach proposed by Nylander *et al*. (2008), which integrates DIVA parsimony-based reconstructions over a Bayesian MCMC sample of trees representing the posterior probability of the tree topology (hereafter referred to as Bayes-DIVA). Specifically, we sampled one tree for every 16 trees (7501 in total) from the MCMC 'post-burnin' sample and used R scripts available from the second author to summarize/average ancestral area reconstructions over all sampled trees for each node in the 50% majority-rule Bayesian inference, which was used as the reference. Only those trees containing the node of interest were summarized in estimating the probabilities for that node. This approach allows an estimation of the marginal probabilities of ancestral ranges for a given node whilst integrating over the uncertainty in the rest of the tree topology (Nylander *et al*., 2008). Ancestral areas and vicariance/dispersal events were recorded following Buerki (2009).

As several polytomies were found in the 50% majority-rule topology, the following rule was applied to solve incompatibilities between nodes and to estimate correct dispersal-vicariance events (that otherwise would violate DIVA assumptions): if the most probable area for a given node was incompatible (according to DIVA assumptions) with that of the next coming node or tip, it was combined with the following most probable area(s); this was performed until the ancestral areas of the node were congruent with the areas assigned to the following node or tip. In order to summarize the different dispersal events across the three geological epochs spanning the diversification of *Arum* (Miocene, Pliocene and Pleistocene), a pairwise matrix of dispersion was built for each epoch to address the links among the ten defined areas [this was performed using R (R Development Core Team, 2009), with scripts available on request from the second author]. When a branch spanned over more than one epoch, the proportion of the branch over each epoch was considered, and the fraction D of one single dispersal event in a given epoch was recorded ($0 < D < 1$). To summarize the results, arrows with variable widths (proportional to the number of dispersal events) were drawn on palaeogeographical maps corresponding to the three relevant epochs (Meulenkamp & Sissingh, 2003).

RESULTS

Phylogenetic inferences

The combined dataset consisted of 250 sequences. Aligned lengths were 845 bp for *3′rps16-5′trnK*, 1077 bp for the *ndhA* intron, 1024 bp for *psbD-trnT*, 777 bp for *rpl32-trnL* and 104 binary positions corresponding to coded gaps. The final matrix thus contained a total of 3827 characters (3723 nucleotides and 104 gap presence/absence). Values for C, V and PI for each partition are given in Table 2. Partition *rpl32-trnL* provides slightly more informative sites (36) than the other partitions. The partition homoge-

neity test was passed ($P = 0.07$), indicating that the information provided by the four plastid regions was congruent. Topologies obtained with Bayesian inference and MP algorithms (Fig. 1) were highly congruent (quartet distance of 0.114, meaning that approximately 89% of the components were compatible between the two trees) and defined five major supported clades (see Fig. 1). Topologies depict *A. pictum* as sister to the rest of the genus, confirming the definition of subgenera *Gymnomesium* and *Arum* (supported by BPP = 1 and DI = 10; Fig. 1). In subgenus *Arum*, the first branching clade I includes *A. palaestinum* Boiss. and the different subspecies of *A. dioscoridis* Sibth. & Sm. (BPP = 1, DI = 3). Clade II contains *A. concinnatum* Schott and *A. italicum* (for which monophyly was not contradicted in the Bayesian analysis) (BPP = 0.95, DI = 3). Clade III includes the two most eastern taxa, *A. jacquemontii* Blume and *A. korolkowii* Regel, *A. megobrebi* Lobin, M.Neumann, Bogner & P.C.Boyce, *A. rupicola* Boiss. and the two easternmost samples of *A. maculatum* L. included in this study, and a new species, hereafter referred to as *A.* sp. nov. (BPP = 0.99, DI = 1). In the MP topology, *A. creticum* Boiss. & Heldr. is also included in clade III as the first branching lineage (Fig. 1). Clade IV includes two subclades containing *A. sintenisii* P.C.Boyce and *A. hygrophilum* Boiss., on the one hand, and *A. byzantinum* Schott, *A. nigrum* Vell. and some specimens of *A. elongatum* Steven, on the other (BPP = 0.84, DI = 1). In the Bayesian inference topology, *A. idaeum* Coust. & Gand. is also included in clade IV as the first branching lineage (Fig. 1). Finally, clade V is poorly resolved and includes the remaining taxa: *A. maculatum* (western samples), *A. cylindraceum* Gasp., *A. orientale* M.Bieb., *A. besserianum* Schott, *A. balansanum* R.R.Mill, *A. purpureospathum* P.C.Boyce, *A. euxinum* R.R.Mill, *A. apulum* (Carano) P.Boyce, *A. cyrenaicum* Hruby and one representative of *A. elongatum* (BPP = 1, DI = 2). The relative position of clade V swapped depending on the phylogenetic algorithm, as it was sister to clade IV in the Bayesian inference tree, but sister to clade III in the MP tree (Fig. 1). Incongruence between the two topologies concerned (1) the positions of clade V relative to clades III and IV, and (2) the branching of *A. idaeum* and *A. creticum*.

Sections as defined by Boyce (1989) were not supported by the phylogenetic hypotheses. Except for the cases of subsections *Poeciloporphyrochiton* (clade I, *A. dioscoridis* and *A. palaestinum*) and *Tenuifila* (subclade in clade III, *A. jacquemontii*, *A. korolkowii* and *A. rupicola*), our topologies did not support the current infrageneric delimitation (Fig. 1, Table 1). Finally, the monophyly of several widespread species (e.g. *A. elongatum* and *A. maculatum*) was not supported (see Fig. 1).

Character evolution

The reconstruction of ancestral states for the four studied characters is shown in Figure 2. The trait that appears to be most constrained from the phylogenetic reconstruction is ploidy (Fig. 2A), with an ancestral character state corresponding to diploidy ($2n = 2x = 28$) and one single evolution towards hexaploidy. Tetraploidy evolved several times. The remaining characters (Fig. 2B, tuber shape; Fig. 2C, flower shape; Fig. 2D, spathe/spadix ratio) show a pattern of multiple independent events and are much less informative at the infrageneric level. An exception could be the evolution of the rhizomatous tuber shape which, although largely symplesiomorphic, seems to be correlated with the level of ploidy.

Biogeographical analysis

Reconstructed ancestral areas for the different nodes corresponding to the 50% majority-rule Bayesian tree are shown in Figure 3. The two most probable ancestral areas from the crown nodes of the genus were the Aegean and West European regions. Later nodes show that the Aegean and Anatolian regions were the only areas to harbour ancestral lineages of the genus for a long time. Overall, a substantial proportion of the dispersion of *Arum* lineages towards their current distribution areas seems to have happened after the late Miocene (*c.* 10 Mya).

The rates and direction of dispersal events at three different time-slices corresponding to the Miocene, Pliocene and Pleistocene are shown in Figure 4. During the Miocene (23–5.3 Mya, Fig. 4A), dispersion mainly happened from the Aegean area to Anatolia. Exchanges were also possible between the newly emerging Caucasus region and the Aegean and Anatolian areas. Colonization of the Iranian area seems to have happened only during this period. Colonization of Macaronesia also occurred at this time, but other dispersals towards this area were probably also possible during the early Pliocene (5.3–2.6 Mya, Fig. 4B). During this epoch, the genus extended its distribution for the first time onto the Arabian plate, and important dispersion events seem to have happened from the Aegean (and, to a lesser extent, from the Anatolian region) to Eastern Europe. The North African region was colonized during the most recent geological epoch (Pleistocene, 2.6 Mya–present, Fig. 4C) probably via two pathways: (1) from the Apennines through the southern tip of the Italian Peninsula; and (2) from the Arabian region through the Gulf of Suez. Numerous dispersals also occurred from the Aegean to the Anatolian area in the Pleistocene and from the Apennine region to Western and Eastern Europe and the Aegean. During this last epoch, exchanges seemed to have halted between the

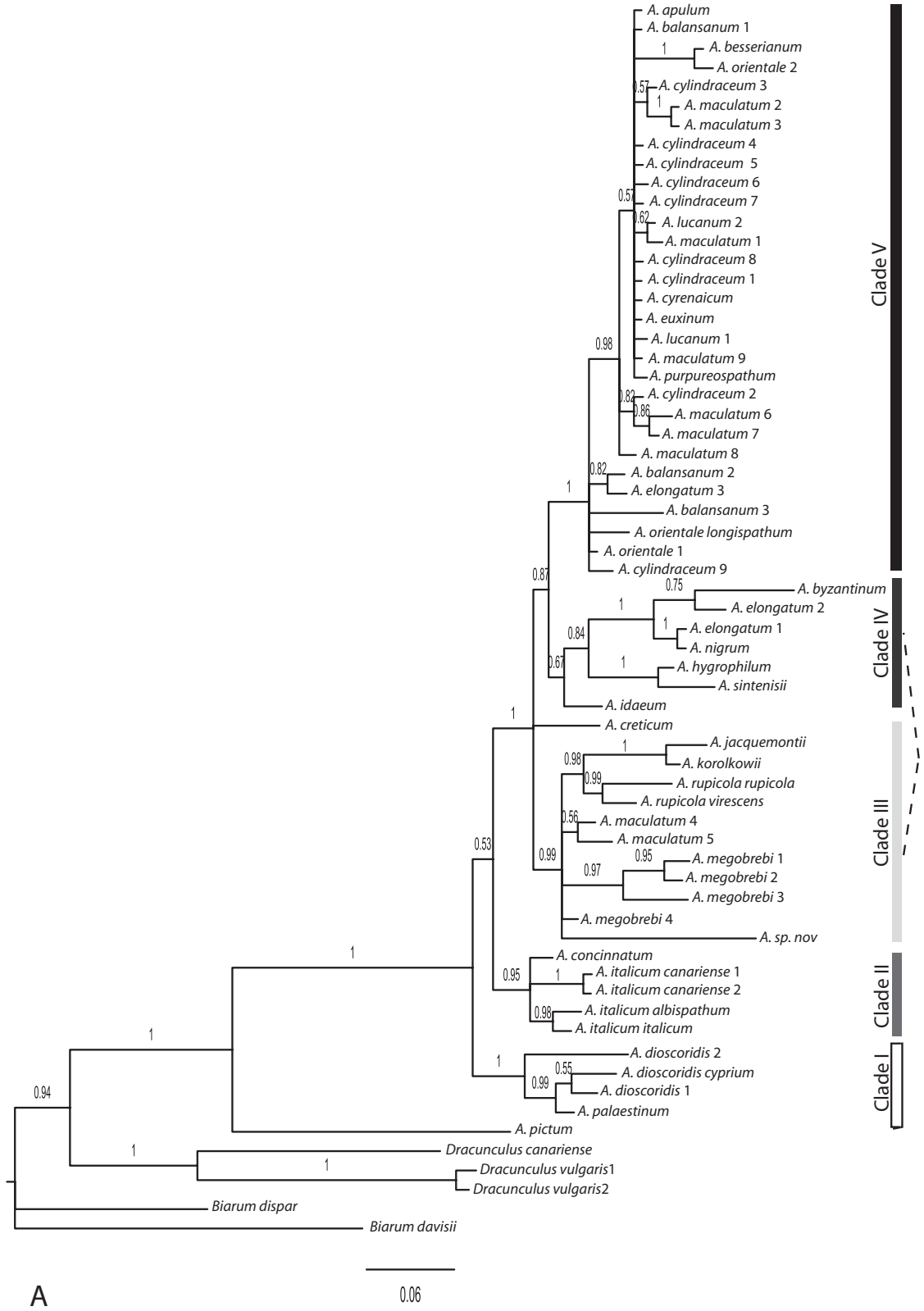


Figure 1. Inferred plastid phylogenies: A, Bayesian inference, half-compatible tree; B, maximum parsimony (MP), strict consensus tree. Values shown on the branches represent Bayesian posterior probability values (A) and decay indices (B). Vertical bars indicate major clades explained in the text.

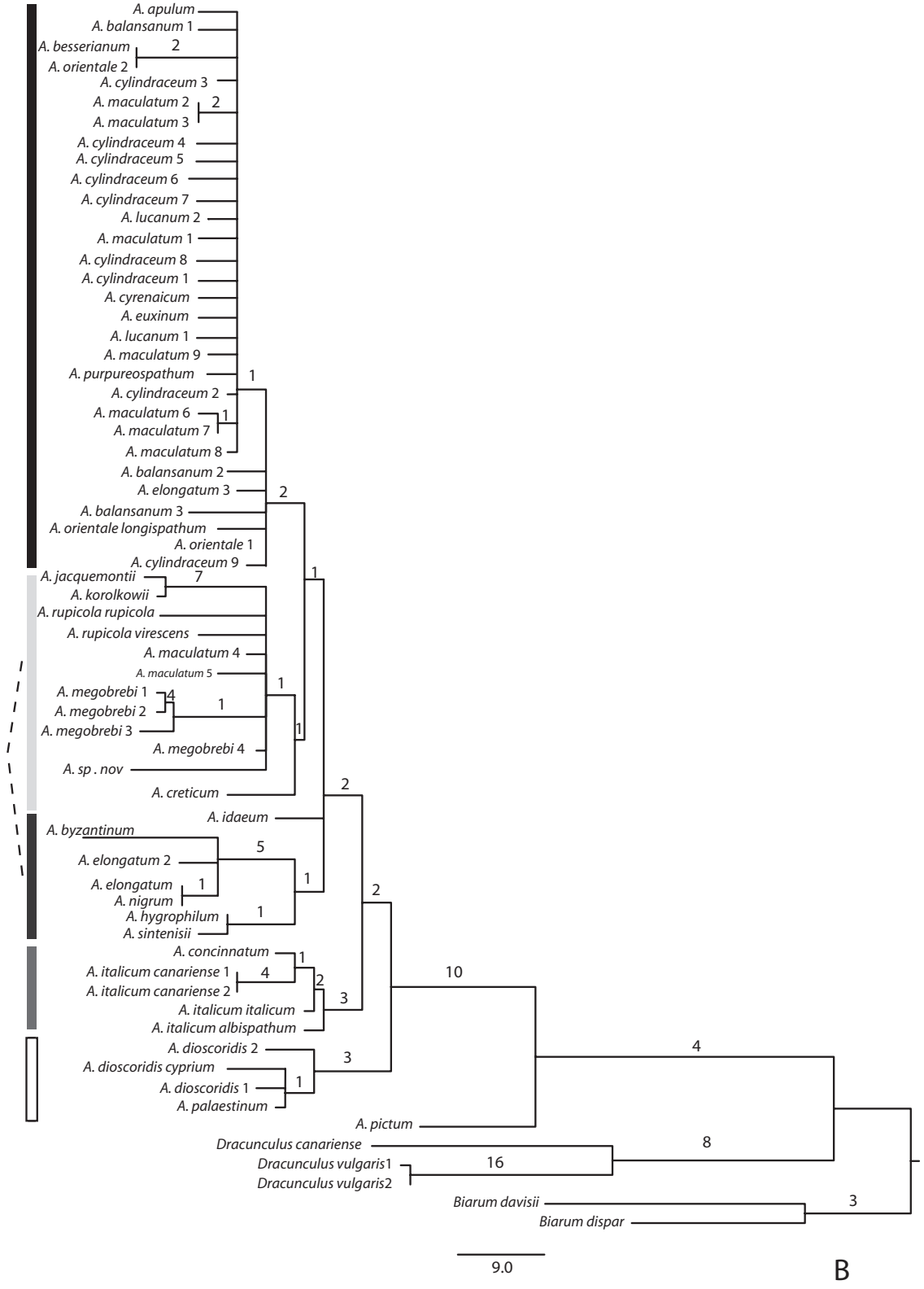


Figure 1. *Continued*

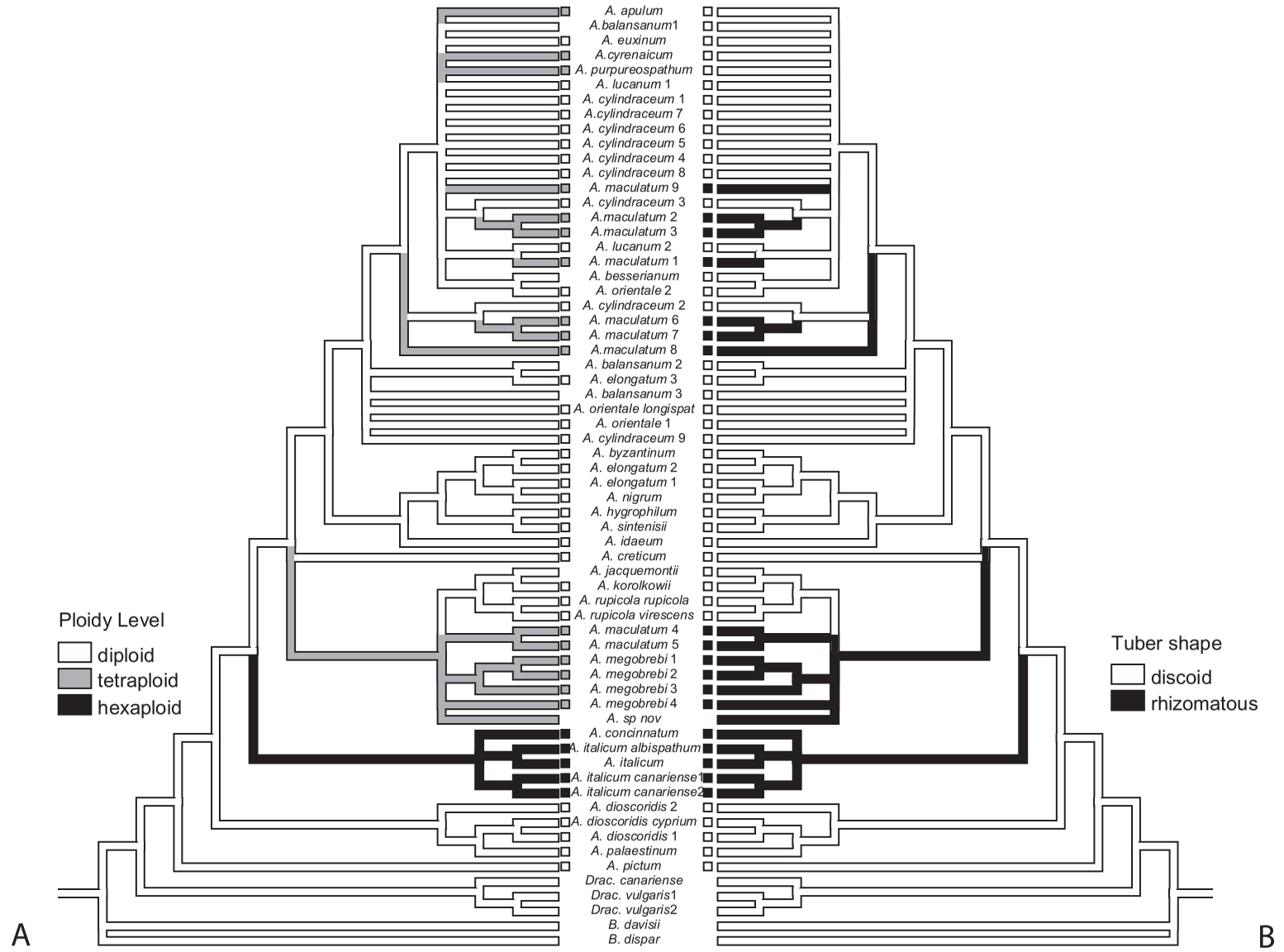


Figure 2. Evolution of categorical characters on the Bayesian topology: A, ploidy ($2x$, $4x$, $6x$); B, tuber shape (rhizomatous, discoid); C, flower type (flag, cryptic); D, spadix/spathe ratio (0–0.5; 0.5–1; > 1). Colours are explained in the legend of each figure. Missing squares at the level of terminal taxa indicate unknown characters.

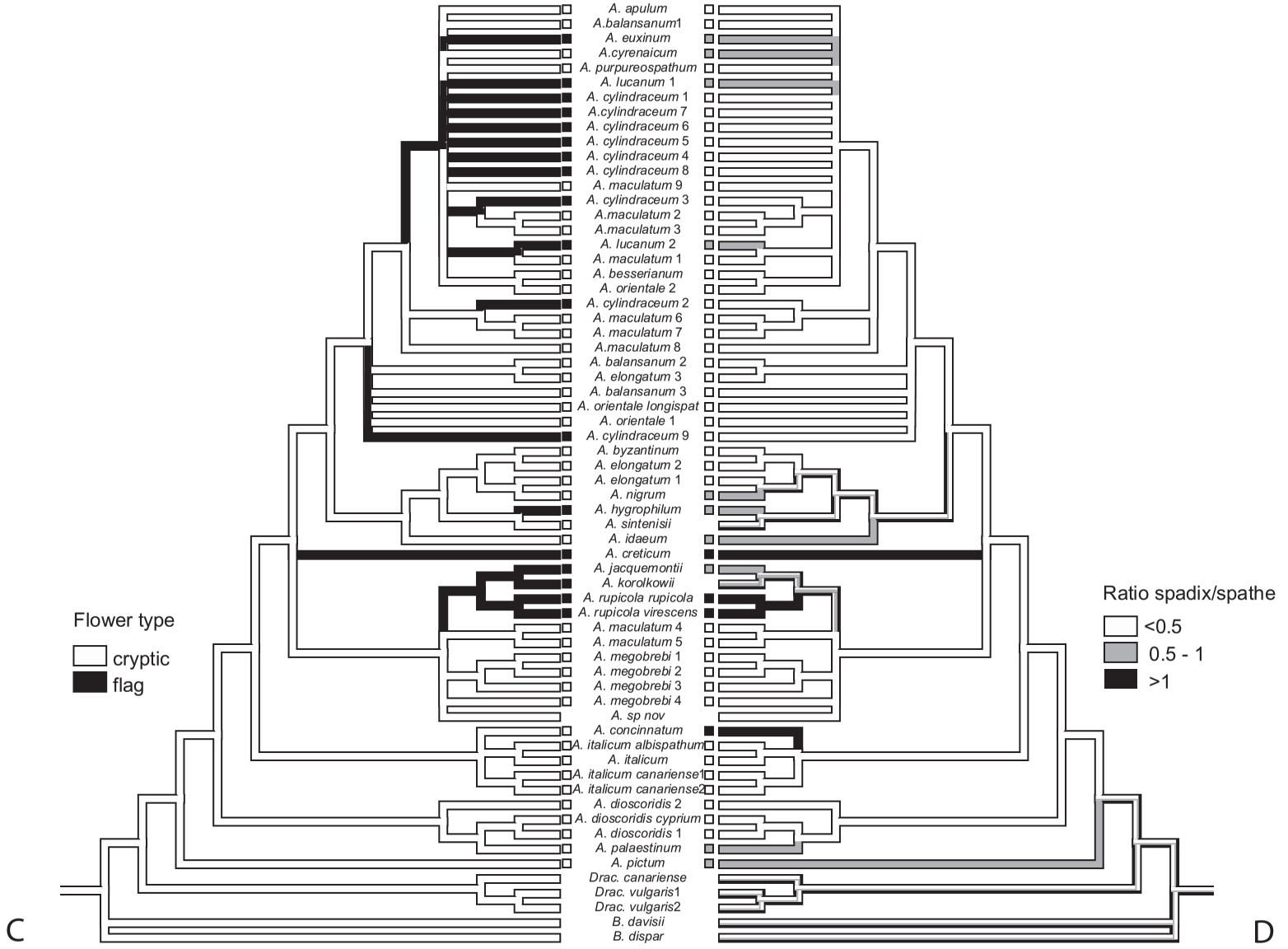


Figure 2. *Continued*

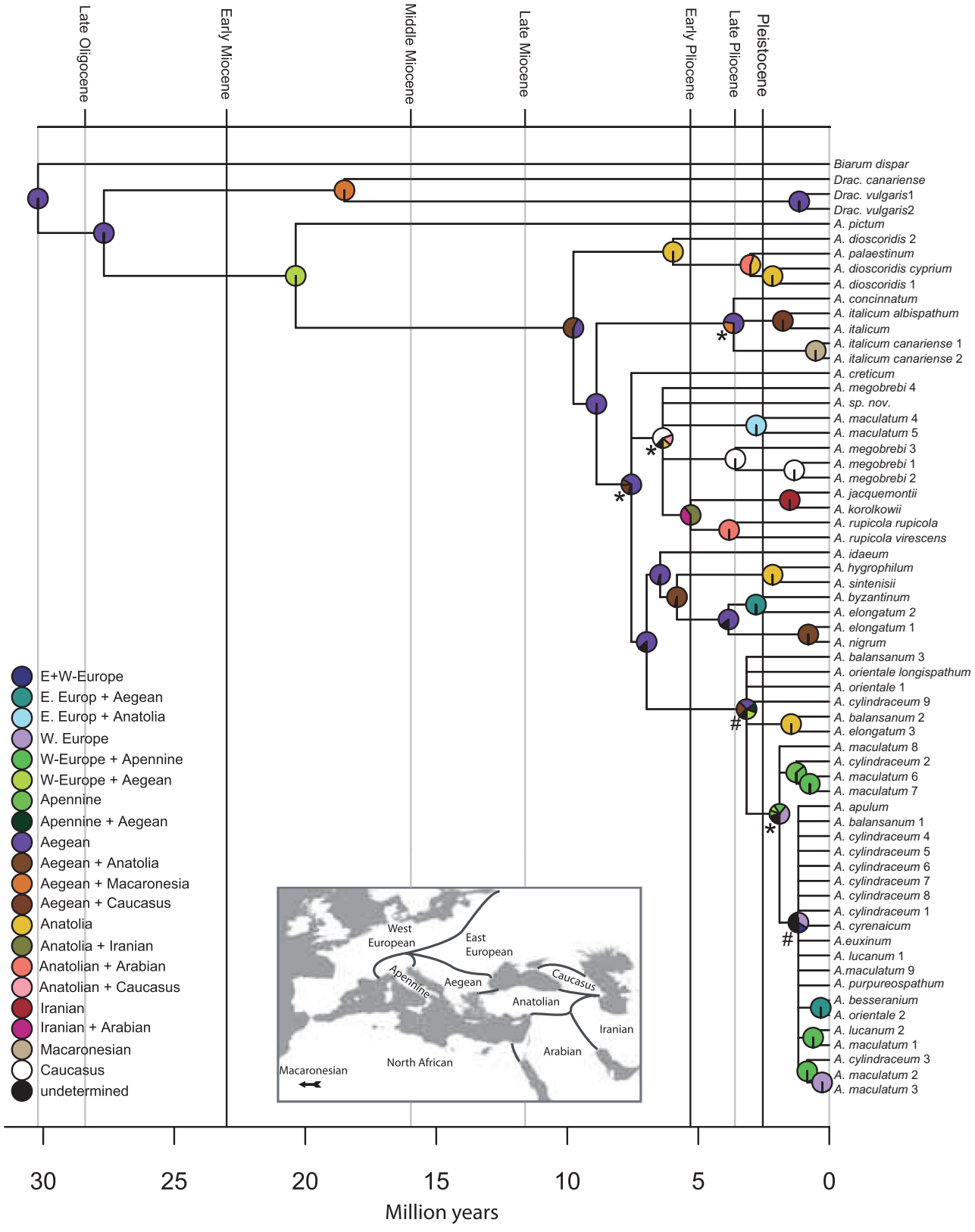

Late Oligocene
Early Miocene
Middle Miocene
Late Miocene
Early Pliocene
Late Pliocene
Pleistocene
Biarum dispar
Drac. canariense
Drac. vulgaris1
Drac. vulgaris2
A. pictum
A. dioscoridis 2
A. palaestinum
A. dioscoridis cyprium
A. dioscoridis 1
A. concinnatum
A. italicum albispathum
A. italicum
A. italicum canariense 1
A. italicum canariense 2
A. creticum
A. megobrebi 4
A. sp. nov.
A. maculatum 4
A. maculatum 5
A. megobrebi 3
A. megobrebi 1
A. megobrebi 2
A. jacquemontii
A. korolkowii
A. rupicola rupicola
A. rupicola virescens
A. idaeum
A. hygrophilum
A. sintenisii
A. byzantinum
A. elongatum 2
A. elongatum 1
A. nigrum
A. balansanum 3
A. orientale longispathum
A. orientale 1
A. cylindraceum 9
A. balansanum 2
A. elongatum 3
A. maculatum 8
A. cylindraceum 2
A. maculatum 6
A. maculatum 7
A. apulum
A. balansanum 1
A. cylindraceum 4
A. cylindraceum 5
A. cylindraceum 6
A. cylindraceum 7
A. cylindraceum 8
A. cylindraceum 1
A. cyrenaicum
A.euxinum
A. lucanum 1
A.maculatum 9
A. purpureospathum
A. besseranium
A. orientale 2
A. lucanum 2
A. maculatum 1
A. cylindraceum 3
A. maculatum 2
A. maculatum 3
E+W-Europe
E. Europ + Aegean
E. Europ + Anatolia
W. Europe
W-Europe + Apennine
W-Europe + Aegean
Apennine
Apennine + Aegean
Aegean
Aegean + Anatolia
Aegean + Macaronesia
Aegean + Caucasus
Anatolia
Anatolia + Iranian
Anatolian + Arabian
Anatolian + Caucasus
Iranian
Iranian + Arabian
Macaronesian
Caucasus
undetermined
West European
East European
Apennine
Aegean
Caucasus
Anatolian
Iranian
Arabian
North African
Macaronesian
30
25
20
15
10
5
0
Million years

Figure 3. Ancestral areas assigned by dispersal-vicariance analysis (DIVA) to each node of the Bayesian topology. Colours represent ancestral areas (see legend). All areas with a probability < 0.1 were pooled and treated as a single undetermined area (in black). Nodes having been treated with special rules are indicated by '*' or '#': '*' indicates that the second most probable area has been combined with the first in order to sketch compatible scenarios; '#' indicates that more than two ancestral areas have been combined to obtain the congruence of the nodes and tips. Scale corresponds to million years from present. Map shows areas defined for the biogeographical inference.

Aegean and Eastern Europe, despite the neighbouring position of these two regions.

DISCUSSION

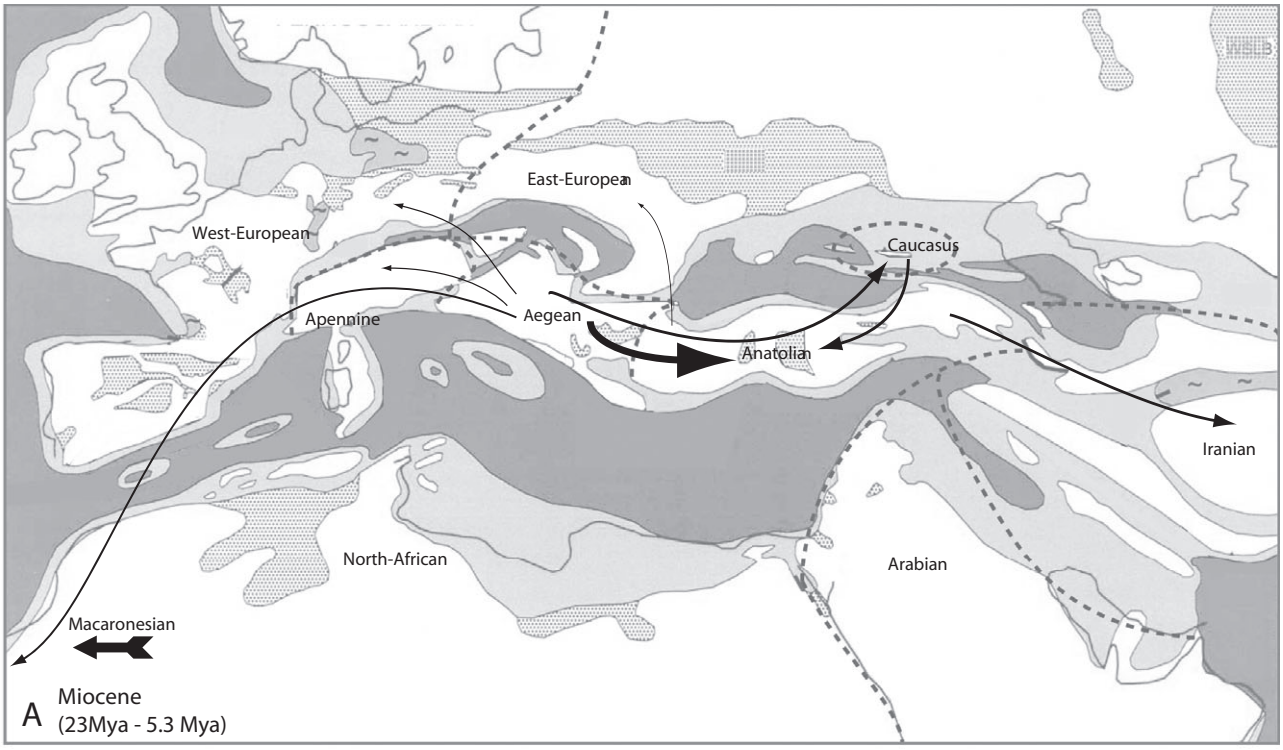
Infrageneric relationships and species' identity

Classical taxon definition and circumscription in the genus *Arum* (Boyce, 1993, 2006) only partially match our phylogenetic hypothesis. As shown in Figure 1, the identity of sections, subsections (*sensu* Boyce, 1989, 1993) and species is strongly challenged, and it seems obvious that there is a 'gap' between the current classification and the genetic identities of the taxa. Our analyses, however, confirm the validity of the two subgenera, as *A. pictum* (subgenus *Gymnomesium*) is the first branching lineage of *Arum* (as shown previously by Mansion *et al*., 2008). This result is supported by floral (staminodes present, but no pistillodes) and phenological (flowering in autumn and not in spring as in the rest of the species) characters restricted to *A. pictum*. For more than a century, the peculiarities of this Hercynian endemic have been recognized, and several authors have attempted to place it in a different genus (*Gymnomesium* Schott, 1855). The long branch separating this lineage from the other species (with a divergence estimated to be sometime between the early and middle Miocene; Fig. 3) favours the hypothesis of a palaeorelictual identity of *A. pictum* (as proposed by Mansion *et al*., 2008). Within subgenus *Arum*, only subsection *Poeciloporphyrochiton* (Fig. 1, clade I) is retrieved by our phylogenetic hypothesis. It appears as the first branching lineage in the subgenus, confirming that *A. dioscoridis* and *A. palaestinum* are closely related and placed in an external position, as proposed formerly by several authors (Hruby, 1910; Boyce, 1989, 1993). Another exception could be subsection *Tenuifila*, which is nested in clade III in the Bayesian topology, and might still be considered as a valid entity (see below). No other subsection is compatible with our results.

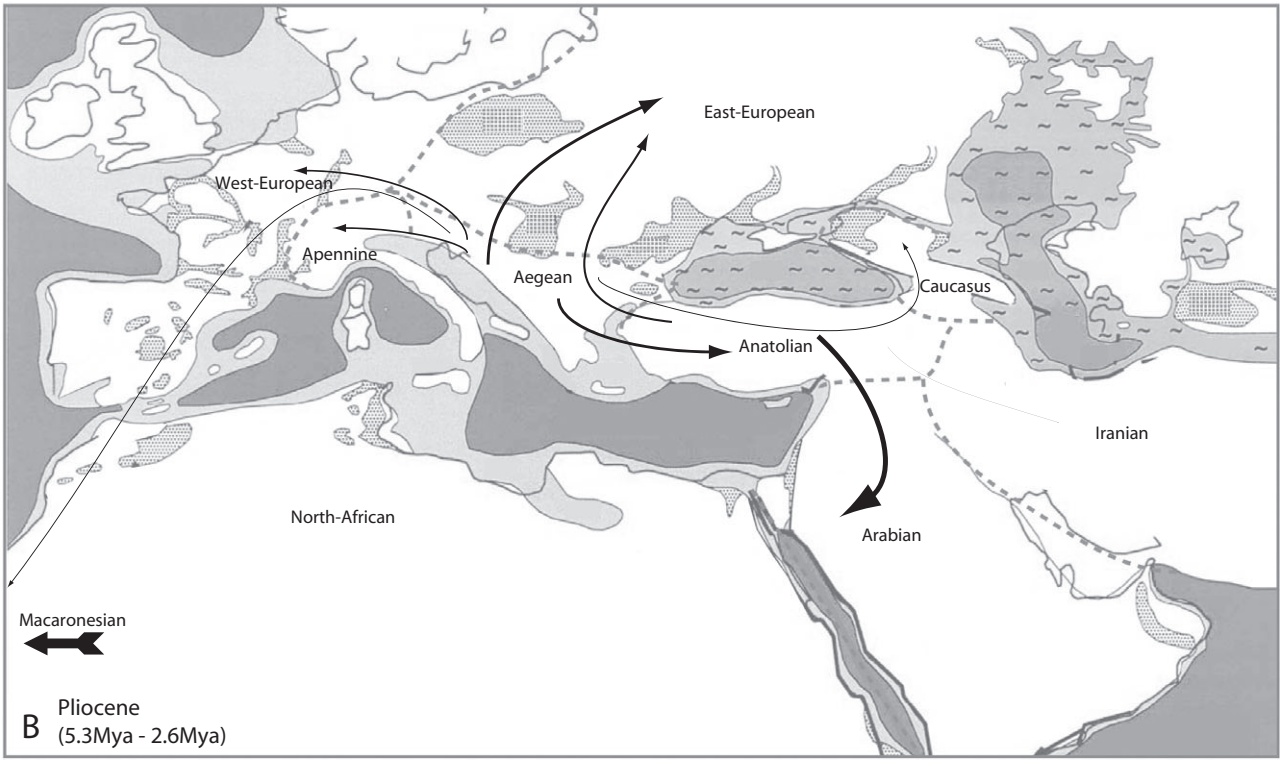
Both Bayesian and MP topologies argue in favour of the monophyly of hexaploid taxa (Fig. 1, clade II; Fig. 2A), with *A. italicum* specimens clustering together with *A. concinnatum*. The insular *A. italicum* ssp. *canariense* (Webb & Berthel.) P.Boyce is genetically differentiated from the 'continental' subspecies from which it diverged during the Pliocene (Fig. 3). As the monophyly of *A. italicum* is not retrieved in the MP analysis, a more thorough analysis (e.g. using genomic screening markers) should be performed to confirm the status of *A. concinnatum*.

Although the phylogenetic relationships among the three remaining clades (III, IV and V) are not yet resolved (i.e. the topology varies according to the phylogenetic algorithm), their respective monophyly is relatively well supported with DI $\geq$ 1 and BPP > 0.95 (with the exception of clade IV, which shows a lower support of 0.84; see Fig. 1). Current molecular data do not allow the discussion of the phylogenetic relationships of *A. creticum* and *A. idaeum*, two species with overlapping distributions in Crete, which are either placed in a polytomy at the base of these three clades or as the first branching lineages of clade III (MP topology) and clade IV (Bayesian topology), respectively. These species are morphologically similar (open floral chambers, sweet or weak odour production vs. closed floral chamber and clear lure-oriented odour production in the other species) and were included in subsection *Cretica* (Boyce, 1989). Relationships among clades III, IV and V should also be examined carefully as our results do not allow conclusions to be drawn regarding the position of clade V, as it appears as either sister to clade III (MP topology) or to clade IV (Bayesian inference tree).

The strongly supported clade III (excluding *A. creticum*; BPP = 0.99; DI = 1) comprises all members of subsection *Tenuifila* (i.e. *A. rupicola*, *A. jacquemontii* and *A. korolkowii*, which form a well-supported monophyletic group in the Bayesian topology; BPP = 0.98) and all representatives of *A. megobrebi*, two specimens of *A. maculatum* from the easternmost edge of the distribution and one sample from the Caucasus area. The latter should be considered as a new species (referred to as *A.* sp. nov.). The placement of *A. maculatum* samples in clade III is unexpected as the two specimens found here are highly divergent phylogenetically with respect to the European representatives (clade V) of this widely distributed species (Fig. 1). As a consequence, they might merit treatment as a different species if further morphological studies confirm this status by identifying synapomorphies. Clade III has a biogeographical coherence as the taxa included are found in the eastern part of the



A Miocene
(23Mya - 5.3 Mya)



B Pliocene
(5.3Mya - 2.6Mya)

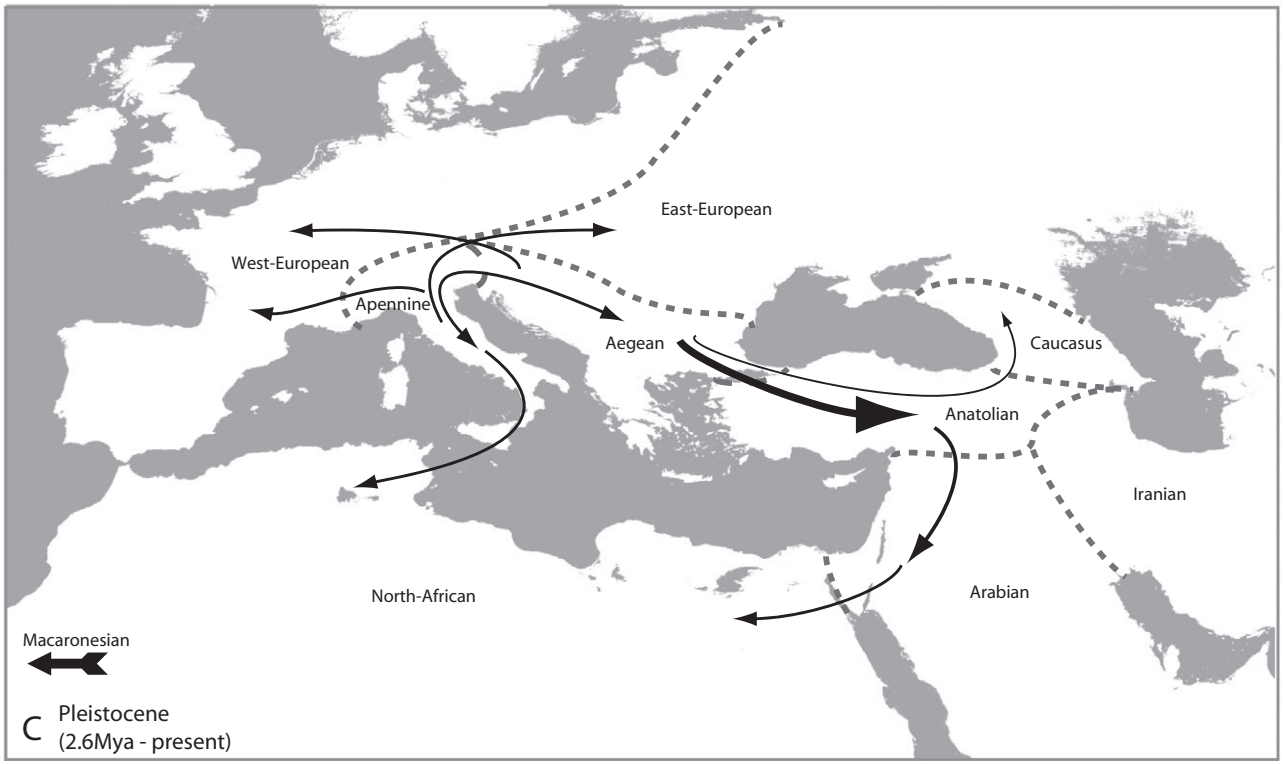


Figure 4. Dispersion events at three time slices: A, Miocene; B, Pliocene; C, Pleistocene (maps A and B, modified from Meulenkamp & Sissingh, 2003; with permission of the editors). The widths of arrows are proportional to dispersal rates. Broken lines indicate boundaries between biogeographical zones.

distribution area of the genus (from northern Turkey to the Iranian region), confirming the role played by this area in the diversification of *Arum* (as proposed previously by Bedalov & Küpfer, 2005).

Clade IV (excluding *A. idaeum*; BPP = 0.84; DI = 1) comprises two specimens of *A. elongatum*, which appears to be paraphyletic with respect to *A. nigrum*, and possibly *A. byzantinum* (in the Bayesian topology), although the position of the latter is not well supported (Fig. 1). More generally, the morphology typical for *A. elongatum* seems to be quite labile as one specimen is also found in clade V (see below). The other representatives of this clade are the closely related *A. sintenisii* (endemic to Cyprus) and the oriental *A. hygrophilum*, the former probably having diverged from the latter in the late Pliocene after a dispersal followed by an insular differentiation. This result was already predicted by Boyce (2006).

Finally, clade V is by far the least resolved, encompassing closely related taxa that diverged during the second half of the Pliocene and the Pleistocene (Fig. 3), most having colonized the Apennines and temperate habitats in Western and Eastern Europe (Fig. 4). In this clade, differentiation among specimens is weak and all species sampled more than once are paraphyletic (Fig. 1). Uncovering the relationships among taxa within this clade would require further analyses based on, for example, genomic screening. This might help to address the status of widely distributed taxa, such as *A. cylindraceum* and the 'European' *A. maculatum*, with that of narrow endemics such as *A. apulum* and *A. purpureospathum*. Another case of interest is the well-supported group composed of *A. besserianum* and one specimen of *A. orientale* (BPP = 1; DI = 2; Fig. 1). *Arum orientale* is still poorly defined as attested by successive revisions during the last 15 years (Boyce, 1994, 2006; Bedalov & Küpfer, 2005). This taxon was first described as a species with several subspecies, present in Crimea and extending to the eastern part of the Balkans (Boyce, 1993). However, several morphological characters point to a close relationship with *A. besserianum* distributed in Ukraine and Poland (P. Küpfer, pers. observ.). Consequently, this taxon certainly encompasses different paraphyletic lineages, and both its status and that of other taxa (e.g. *A. balansanum*) should be investigated using novel genomic techniques coupled with taxonomy.

Therefore, we have observed that two different patterns arise when testing the monophyly of species for which more than one specimen was collected. On the one hand, some are well supported by our analyses:

this is the case for *A. dioscoridis* (clade I), *A. rupicola* (clade III) and, to a lesser extent, *A. megobrebi* (clade III) (Fig. 1). On the other hand, some species are clearly polyphyletic (comprising specimens from lineages that diverged as early as the Pliocene), such as, for instance, *A. maculatum* (clades III and V) and *A. elongatum* (clades IV and V) (Fig. 1). Finally, the case of *A. italicum* (clade II) is somewhat intermediate, as the monophyly is not contradicted by the Bayesian topology, whereas the species is paraphyletic (with the inclusion of *A. concinnatum*) in the MP topology.

Character evolution: characterizing the identity of the most recent common ancestor

Among all the investigated characters, only ploidy seems to be related to the evolution of the genus (Fig. 2A). The remaining traits (spathe/spadix ratio, flower type, shape of the tuber) show patterns of multiple independent evolution and a looser correlation with the evolutionary history of *Arum*. Our results therefore support the hypothesis that the level of ploidy might constitute an informative character for the systematics of the genus, as first proposed by Bedalov & Küpfer (2005), and address the diploid status of the most recent common ancestor (a trait shared by both *A. pictum* and taxa within clade I). However, the abrupt transition from diploidy to hexaploidy (in clade II) seems to be unlikely and might require the existence of a transitional and yet extinct or undiscovered tetraploid form. The advantages of polyploids in terms of survival have been addressed recently in *Arum*, as artificial crossings between distinct species yielded polyploid hybrids that were 'robust and maintain themselves in cultivation without apparent difficulties' (Bedalov & Küpfer, 2005). Interestingly, the fact that clade V comprises both a substantial proportion of tetraploid lineages and an important number of recognized taxa could support the idea of an increased fitness in polyploids, facilitative for the radiation of this group (Fig. 3) (for a review of the ability of polyploids to colonize a wider range of habitats, see Prentis *et al*., 2008). However, before arriving at any conclusion, it is important that the phylogenetic relationships among the specimens of this clade are clarified.

One local phylogenetic constraint on the flower type (flag vs. cryptic) was addressed in the subclade corresponding to the *Tenuifila* subsection, with all taxa sharing a flag flower, whereas the ancestral state for this trait within *Arum* was a cryptic flower. As there is a strong association between cryptic flowers and attract-and-lure pollination strategies (Boyce, 1989; Gibernau *et al*., 2004), the latter should be considered as the ancestral pollination mode in the genus. It is important, however, to mention that the polyphyletic status of this character is not surprising, as it is related to reproductive structures, which, in *Arum*, appear to be highly correlated with fast-evolving pollination syndromes (Chouteau, Gibernau & Barabé, 2008). The evolution of this character would thus reflect more strongly the ecological processes that species have independently undergone rather than the evolutionary history of the genus.

Finally, our results confirm that the ancestral state of the *Arum* tuber shape was discoid, as proposed by Bedalov & Küpfer (2005), and that the appearance of the first rhizomatous species happened at the same time as the transition from diploidy to hexaploidy (clade II, Fig. 1). Although the transition from a discoid to a rhizomatous tuber occurred several times, there seems to be a trend towards a correlation between ploidy and tuber shape: all rhizomatous species are polyploid. In contrast, several polyploid species (*A. apulum*, *A. cyrenaicum* and *A. purpureospathum*) have discoid tubers. As sections within *Arum* were classically defined on the basis of this homoplasious character, there is a strong need to consider morphological characters from other plant parts to build a new classification of *Arum* compatible with our molecular evidence.

Arum through space and time

Bayes-DIVA provides strong support for an Aegean/Western European origin of the genus (Fig. 4) sometime in the early Miocene (*c.* 20 Mya). However, assuming that the earliest diverging lineage (now composed of only *A. pictum*) originated and survived in the Hercynian islands long before all other *Arum* spp. arose (according to the palaeorelictual hypothesis proposed by Mansion *et al*., 2008), the ancestral area corresponding to the rest of the genus is the Aegean region (Fig. 3). This zone has acted as a natural laboratory allowing the diversification of lineages, sometime in the late Miocene (Figs. 3, 4). The Aegean also appears to be a main source of dispersal events throughout the evolutionary history of *Arum*. Its central position with respect to the other areas in which the genus is present today could have facilitated this. Most dispersal events recorded during the middle to late Miocene occurred from the Aegean to the Anatolian region (Fig. 4A). Later, the emergence of the Iranian plate allowed its colonization once a landbridge was established with the Anatolian plate in the late Miocene (Meulenkamp *et al*., 2000). During this period, no dispersals were observed towards the Arabian plate that was still isolated from the northern lands by a marine transgression (Meulenkamp & Sissingh, 2003). Once the Caucasian archipelago emerged (and possibly after its uplift and contact with the Northern Anatolian region), further dispersals to

and from this region occurred in the late Miocene. At this same period, long-distance dispersals are recorded from the Aegean to the Macaronesian regions.

The first dispersal to the Arabian zone occurred more recently, during the Pliocene, when a sea regression (Meulenkamp & Sissingh, 2003) allowed this land to come into contact with the Anatolian region (Fig. 4B). The regression of the western Para-Tethys (following the uplift of the European plates) could also have permitted the dispersal from the Aegean to Eastern Europe. Exchanges continued between the Aegean and Anatolia and through these two zones to the Caucasus and Eastern Europe.

During more recent times (Pleistocene), new dispersals from the Aegean to Anatolia were recorded (Fig. 4C), probably facilitated by the Mediterranean regressions characterizing Quaternary climatic oscillations (Peulvast *et al*., 2000). At this time, North Africa was colonized twice through the Arabian plate and through the Apennines (Fig. 4C).

Although exchanges mainly occurred longitudinally (east–west) across land paths at the periphery of the seas during the early history of the genus, the pace of dispersion and diversification in *Arum* increased after the peri-Mediterranean region was unified (i.e. after the emergence of the Arabian and Iranian plates, the uplift of the Caucasus and the regressions of Tethys and Para-Tethys) (Meulenkamp *et al*., 2000; Meulenkamp & Sissingh, 2003).

The colonization of Macaronesia requires special treatment. The taxon inhabiting this region (*A. italicum* ssp. *canariense*) appears to have arrived there during the late Miocene or early Pliocene, in agreement with the timing of colonization already observed in several other endemic taxa of these islands (Carine *et al*., 2004). This ancient dispersal contrasts with the more recent colonization of North Africa. This discordance could be a result of either a first colonization of North Africa, having allowed the dispersal towards Macaronesia through mid-distance dispersal and further extinction of this lineage, or a long-distance dispersal directly from the Northern Peri-Tethys. Considering the morphology of the seeds (Mayo *et al*., 1997), this latter hypothesis could be possible only in association with animals; birds have already been proposed as the main dispersers of *A. maculatum* (Snow & Snow, 1988), which could also be true for *A. italicum* (Méndez, 1997).

Towards a new classification of *Arum*?

Because our phylogenetic reconstruction strongly contradicts the current systematics of the genus, the need for a new classification is evident (i.e. a large number of homoplasies are suggested by the tracing of the characters currently used in the delimitation of sections and species in our topology; Fig. 2). However, we recommend caution in formally proposing a new infrageneric classification until nonmolecular synapomorphies supporting the main clades are identified.

The two current subgenera, *Arum* and *Gymnomesium*, are supported by our analyses, although the status of the latter might be reconsidered, given the high level of phylogenetic differentiation of this monospecific subgenus. The characteristic morphology, development, distribution and, as shown in this study, phylogenetic position displayed by *A. pictum* could indicate that it would be more correct to place it in the monospecific genus *Gymnomesium* (as formerly proposed by Schott, 1855).

The two formerly defined sections within subgenus *Arum* are not supported by the phylogenetic analyses and, based on molecular evidence, we recommend a division of the subgenus into five sections (corresponding to clades I–V). Subsection *Poeciloporphyochiton* (corresponding to clade I) should be elevated to the rank of section, whereas new synapomorphies should be recovered for the other clades. The sectional classification of *A. creticum* and *A. idaeum* should also be investigated more thoroughly as our phylogenetic hypotheses only weakly associate them with clades III and IV, respectively. Although this study demonstrates the importance of ploidy as a putative synapomorphy in the case of clade II, a broad survey of morphological characters is strongly recommended.

Finally, the paraphyletic status of widespread species requires additional analyses to be performed with more variable markers in order to validate these findings. Nonetheless, our results already argue for a revision of species such as *A. maculatum* and *A. elongatum* in which major splits have been identified. Future taxonomic revisions should carefully consider characters not related to pollination, as lineages seem to be able to adapt quickly to changes in pollinator availability, leading to floral character convergence in distinct clades.

ACKNOWLEDGEMENTS

The authors would like to thank the different persons and institutions that kindly provided the material used in this study: the DNA Bank of the Royal Botanic Gardens, Kew (UK), as well as Dr Cusimano (University of Münich, Germany), Dr Lobim and Mr Neumann (Botanical Garden, University of Bonn, Germany). Discussions with Dr Bogner (Botanical Garden Munich, Germany) also largely contributed to the design of the manuscript. This study was funded by the Swiss National Science Foundation (project no. 3100A0-116778) and the Swiss Academy of Sciences (SCNAT+).

REFERENCES

Akaike H. 1973. Information theory and an extension of the maximum likelihood principle. In: Petrox BN, Caski F, eds. *2nd International Symposium on Information Theory*. Budapest: Akademiai Kiadó, 267–281.

Anderson C, Janssen T. 2009. Monocots. In: Hedges SB, Kumar S, eds. *The timetree of life*. Oxford: Oxford University Press, 203–212.

Baker RH, DeSalle R. 1997. Multiple sources of character information and the phylogeny of Hawaiian drosophilids. *Systematic Biology* **46:** 654–673.

Bedalov M. 1981. Cytotaxonomy of the genus *Arum* (Araceae) in the Balkans and the Aegean area. *Botanische Jahrbücher für Systematik, Pflanzengeschichte und Pflanzengeographie* **102:** 183–200.

Bedalov M, Küpfer P. 2005. Studies on the genus *Arum* (Araceae). *Bulletin de la Société Neuchâteloise des Sciences Naturelles* **128:** 43–70.

Borchsenius F. 2009. *Fastgap 1.2*. Distributed by the author. Aarhus, Denmark: Department of Biological Sciences, University of Aarhus. Available at http://192.38.46.42/aubot/fb/FastGap_home.htm (accessed 4 February 2010).

Boyce PC. 1989. A new classification of *Arum* with keys to the infrageneric taxa. *Kew Bulletin* **44:** 383–395.

Boyce PC. 1993. The genus *Arum.* A Kew magazine monograph. Royal Botanic Gardens, Kew, London. Norwich: HMSO Books.

Boyce PC. 1994. The *Arum* species of Greece and Cyprus. *Annales Musei Goulandris* **9:** 27–38.

Boyce PC. 2006. *Arum* – a decade of change. *Aroideana* **29:** 132–139.

Bremer K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* **42:** 795–803.

Buerki S. 2009. Worldwide biogeography and systematics of Sapindaceae: a molecular and taxonomic survey combining large data sets and novel methodological approaches. Unpublished DPhil Thesis, University of Neuchâtel, Switzerland.

Cabrera LI, Salazar GA, Chase MW, Mayo SJ, Bogner J, Dávila P. 2008. Phylogenetic relationships of aroids and duckweeds (Araceae) inferred from coding and noncoding plastid DNA. *American Journal of Botany* **95:** 1153–1165.

Carine MA, Russell SJ, Santos-Guerra A, Francisco-Ortega J. 2004. Relationships of the Macaronesian and Mediterranean floras: molecular evidence for multiple colonizations into Macaronesia and back-colonization of the continent in *Convolvulus* (Convolvulaceae). *American Journal of Botany* **91:** 1070–1085.

CATE project. 2010. Creating a taxonomic science – January 2010; continuously updated. Available at http://www.cate-project.org (accessed 4 February 2010).

Chase MW, Fay MF, Devey DS, Maurin O, Rønsted N, Davies TJ, Pillon Y, Petersen G, Seberg O, Tamura MN, Asmussen CB, Hilu K, Borsch T, Davis JI, Stevenson DW, Pires JC, Givnish TJ, Sytsma KJ, McPherson MA, Graham SW, Rai HS. 2006. Multigene analyses of monocot relationships: a summary. *Aliso* **22:** 63–75.

Chouteau M, Gibernau M, Barabé D. 2008. Relationships between floral characters, pollination mechanisms, life forms and habitats in Araceae. *Botanical Journal of the Linnean Society* **156:** 29–42.

Estabrook GF. 1992. Evaluating undirected positional congruence of individual taxa between two estimates of the phylogenetic tree for a group of taxa. *Systematic Biology* **41:** 172–177.

Farris JS, Källersjö M, Kluge AG, Bult C. 1994. Testing significance of congruence. *Cladistics* **10:** 315–319.

Flora Europaea. 2009. *Flora Europaea database*. Edinburgh, UK: Royal Botanical Gardens Edinburgh. Available at http://rbg-web2.rbge.org.uk/FE/fe.html

Fuchs L. 1542. *Historia Stirpium*.

Gelman A, Rubin DB. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* **7:** 457–472.

Gibernau M, Macquart D, Przetak G. 2004. Pollination in the genus *Arum* – a review. *Aroideana* **27:** 148–166.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41:** 95–98.

Hruby J. 1910. Le genre *Arum*. Aperçu systématique avec considérations spéciales sur les relations phylogénétiques des formes. *Bulletin de la Société Botanique de Genève* **4:** 330–371.

Kluge AG. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Systematic Zoology* **38:** 7–25.

Linnaeus C. 1753. *Species plantarum 2*. Stockholm: 964–968.

Lobin W, Neumann M, Bogner J, Boyce PC. 2007. A new *Arum* species (Areae, Araceae) from NE Turkey and Georgia. *Willdenowia* **37:** 445–449.

Maddison WP, Maddison DR. 2009. *Mesquite: a modular system for evolutionary analysis. Version 2.6*. Distributed by the authors. Vancouver, BC, Corvallis, OR: University of British Columbia and Oregon State University. Available at http://mesquiteproject.org (accessed 4 February 2010).

Mansion G, Rosenbaum G, Schoenenberger N, Bacchetta G, Rosselló JA, Conti E. 2008. Phylogenetic analysis informed by geological history supports multiple, sequential invasions of the Mediterranean basin by the angiosperm family Araceae. *Systematic Biology* **57:** 269–285.

Mayo SJ, Bogner J, Boyce PC. 1997. *The genera of Araceae*. Kew: Royal Botanical Garden.

Médail F, Diadema K. 2009. Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography* **36:** 1–19.

Méndez M. 1997. Sources of variation in seed mass in *Arum italicum* (Araceae). *International Journal of Plant Sciences* **158:** 298–305.

Meulenkamp JE, Sissingh W. 2003. Tertiary palaeogeography and tectonostratigraphic evolution of the Northern and Southern Peri-Tethys platforms and the intermediate domains of the African–Eurasian convergent plate boundary zone. *Palaeogeography, Palaeoclimatology, Palaeoecology* **196:** 209–228.

Meulenkamp JE, Sissingh W, Calvo JP, Daams R, Londeix L, Cahuzak B, Kovac M, Marunteanu M, Ilynia LB, Khondkarian SO, Scherba IG, Roger J, Platel J-P, Hirsch F, Sadek A, Abdel-Gawad GI, Yeddi RS, Yaich C, Bouaziz S. 2000. Late Tortonian (8.4–7.2 Ma). In: Crasquin, S (coord.), *Atlas Peri-Tethys. Palaeogeographical maps. Explanatory notes*. Paris: CCGM-CGMW, 195–201.

Minorsky PV. 2003. The hot and the classic. *Plant Physiology* **132:** 1779–1780.

Moodie GEE. 1975. Heat production and pollination in Araceae. *Canadian Journal of Botany* **54:** 545–546.

Nixon KC. 1999. The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics* **15:** 407–414.

Nylander JAA. 2004. *Mraic.pl*. Distributed by the author. Uppsala, Sweden: Evolutionary Biology Centre, Uppsala University. Available at http://www.abc.se/~nylander (accessed 4 February 2010).

Nylander JAA, Olsson U, Alström P, Sanmartín I. 2008. Accounting for phylogenetic uncertainty in biogeography: a Bayesian approach to dispersal-vicariance analysis of the thrushes (Aves: *Turdus*). *Systematic Biology* **57:** 257–268.

Peulvast J-P, Vergne V, Vanney J-R, Barrier E. 2000. Last Glacial Maximum (late Pleistocene, 22000–18000 years BP). In: Crasquin, S (coord.), *Atlas Peri-Tethys. Palaeogeographical maps. Explanatory notes*. Paris: CCGM-CGMW, 209–219.

Prentis PJ, Wilson JRU, Dormontt EE, Richardson DM, Lowe AJ. 2008. Adaptive evolution in invasive species. *Trends in Plant Science* **13:** 288–294.

Prime CT. 1960. *Lords and ladies*. London: Collins.

R Development Core Team. 2009. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rambaut A, Drummond AJ. 2004. *Tracer v1.3*. Distributed by the authors. Edinburgh, UK: Institute of Evolutionary Biology, University of Edinburgh. Available at http://tree.bio.ed.ac.uk/software/tracer (accessed 4 February 2010).

Ree RH, Sanmartín I. 2009. Prospects and challenges for parametric models in historical biogeographical inference. *Journal of Biogeography* **36:** 1211–1220.

Ronquist F. 1997. Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology* **46:** 195–203.

Ronquist F. 2001. *Diva 1.2*. Distributed by the author. Tallahassee, FL: School of Computational Science, Florida State University. Available at http://www.ebc.uu.se/systzoo/research/diva/diva.html (accessed 4 February 2010).

Ronquist F, Huelsenbeck JP. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19:** 1572–1574.

Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution* **19:** 101–109.

Sanderson MJ. 2004. *R8s, program and documentation*. Distributed by the author. Tucson, AZ: Department of Ecology and Evolutionary Biology, University of Arizona. Available at http://loco.biosci.arizona.edu/r8s (accessed 4 February 2010).

Sanmartín I, Enghoff H, Ronquist F. 2001. Patterns of animal dispersal, vicariance and diversification in the Holarctic. *Biological Journal of the Linnean Society* **73:** 345–390.

Schott HW. 1832. Arum. In: Schott HW, Endlicher SL, eds. *Meletemata botanica. Arch. Bot. (Paris)* **1:** 246–252.

Schott HW. 1855. Gymnomesium. *Oesterreichische Botanische Wochenblatt* **5:** 17.

Shaw J, Lickey EB, Schilling EE, Small RL. 2007. Comparison of whole chloroplastic genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany* **94:** 275–288.

Sikes DS, Lewis PO. 2001. *PAUPRat, Beta software, version 1: PAUP implementation of the parsimony ratchet*. Distributed by the authors. Storrs, CT: Department of Ecology and Evolutionary Biology, University of Connecticut. Available at http://users.iab.uaf.edu/~derek_sikes/software2.htm (accessed 4 February 2010).

Simmons MP, Ochoterena H. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* **49:** 369–381.

Snow BK, Snow DW. 1988. *Birds and berries*. Calton, UK: T. and A.D. Poyser.

Sorenson MD, Franzosa EA. 2007. *Treerot, version 3*. Distributed by the authors. Boston, MA: Boston University, Available at http://people.bu.edu/msoren/TreeRot.html (accessed 4 February 2010).

Swofford DL. 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods*)*. Version 4b10. Sunderland, MA: Sinauer Associates.

Theophrastus. 370 B.C. *The enquiry into plants*. English translation by Sir A. Hort, 1916. London: W. Heinemann.

NOTE ADDED IN PROOF

The authors of this study attest that at the time of the acceptance of our manuscript in *Botanical Journal of the Linnean Society* no phylogenetic studies of the genus *Arum* (Araceae) had been published. In the meantime, Linz *et al.* have published a phylogeny of *Arum* in the latest issue of *Taxon* (59: 405–415), i.e. two months before the release of our study. These studies were conducted independently and, although sharing the same group of interest, the sampling, approach and conclusions are consequently slightly different. For the sake of the understanding of the evolution of *Arum*, we encourage readers to compare the present study with that of Linz *et al.*

Linz J, Stökl J, Urru I, Krügel T, Stensmyr MC, Hansson BS. 2010. Molecular phylogeny of the genus *Arum* (Araceae) inferred from multi-locus sequence data and AFLPs. *Taxon* **59:** 405–415.

APPENDIX 1

Sample sources and GenBank accession numbers for all analysed samples. BG-Basle, Botanical Gardens, Basle, Switzerland; RBG-Kew, Royal Botanic Gardens, Kew, UK; pr. coll., private collection; NA, not available. Source information is provided in the following order: collector name, collection, voucher reference.

Sample	Source	Country of origin	GenBank accessions			
			3rps16-trnK	*ndhA*	*psbD-trnT*	*rpl32-trnL*
Arum apulum	Chase, RBG-Kew, 11022K	Italy	GU370965	GU371038	GU371101	GU371165
Arum balansanum1	Chase, RBG-Kew, 11009K	Turkey	GU370966	GU371049	GU371112	GU371174
Arum balansanum2	Küpfer, pr. coll., 09.06.01 4	Turkey	GU370967	GU371065	GU371128	GU371189
Arum balansanum3	Koenen, Haller, pr. coll., NA	Turkey	GU370968	GU371074	GU371138	GU371198
Arum besserianum	Bedalov, pr. coll., 1507	Ukraine	GU370969	GU371061	GU371124	GU371186
Arum byzantinum	Küpfer, pr. coll., 09.06.02 1	Turkey	GU370970	GU371063	GU371126	GU371188
Arum concinnatum	Chase, RBG-Kew, 11014K	Crete	GU370973	GU371060	GU371123	GU371185
Arum creticum	Chase, RBG-Kew, 11037K	Crete	GU370974	GU371068	GU371132	GU371193
Arum cylindraceum1	Chase, RBG-Kew, 11013K	Romania	GU370982	GU371028	GU371091	GU371155
Arum cylindraceum2	Espíndola & Revel, pr. coll., cylGIO1	Italy	GU370975	GU371080	GU371144	GU371204
Arum cylindraceum3	Espíndola & Revel, pr. coll., cylGYL3	Denmark	GU370976	GU371081	GU371145	GU371205
Arum cylindraceum4	Espíndola & Revel, pr. coll., cylMUS4	Poland	GU370977	GU371082	GU371146	GU371206
Arum cylindraceum5	Espíndola & Revel, pr. coll., cylPAD5	Romania	GU370978	GU371083	GU371147	GU371207
Arum cylindraceum6	Espíndola & Revel, pr. coll., cylSIS5	France	GU370979	GU371084	GU371148	GU371208
Arum cylindraceum7	Espíndola & Revel, pr. coll., cylTRA4	Austria	GU370980	GU371085	GU371149	GU371209
Arum cylindraceum8	Espíndola & Revel, pr. coll., cylTRP3	Macedonia	GU370981	GU371086	GU371150	GU371210
Arum cylindraceum9	Neumann, pr. coll., I21/05	Sicily	GU371026	GU371073	GU371137	–
Arum cyrenaicum	Chase, RBG-Kew, 11030K	Libya	GU370983	GU371072	GU371136	GU371197
Arum dioscoridis ssp. *cyprium*	Chase, RBG-Kew, 11021K	Turkey	GU370985	GU371077	GU371141	GU371201
Arum dioscoridis1	Chase, RBG-Kew, 11015K	Turkey	GU370986	GU371078	GU371142	GU371202
Arum dioscoridis2	Küpfer, pr. coll., 09.05.30 4	Turkey	GU370984	GU371064	GU371127	–
Arum elongatum1	Boyce, RBG-Kew, 1990-2019	Turkey	GU370989	GU371079	GU371143	GU371203
Arum elongatum2	Bedalov, pr. coll., 16.5.03	Ukraine	GU370990	–	GU371131	GU371192
Arum elongatum3	Küpfer, pr. coll., 09.05.09 1	Turkey	GU370991	GU371069	GU371133	GU371194
Arum euxinum	Chase, RBG-Kew, 11019K	Turkey	GU370992	GU371029	GU371092	GU371156
Arum hygrophilum	Chase, RBG-Kew, 11027K	Israel	GU370993	GU371030	GU371093	GU371157
Arum idaeum	Boyce, RBG-Kew, 1993-1895	Crete	GU370994	GU371031	GU371094	GU371158
Arum italicum ssp. *albispathum*	Chase, RBG-Kew, 11020K	Georgia	GU370995	GU371035	GU371098	GU371162
Arum italicum ssp. *canariense1*	Chase, RBG-Kew, 11032K	Madeira	GU370996	GU371032	GU371095	GU371159
Arum italicum ssp. *canariense2*	Chase, RBG-Kew, 11031K	Madeira	GU370997	GU371034	GU371097	GU371161
Arum italicum ssp. *italicum*	Boyce, RBG-Kew, 1978-4984	Greece	GU370998	GU371033	GU371096	GU371160
Arum jacquemontii	Boyce, RBG-Kew, 1969-5385	Afghanistan	GU370999	GU371036	GU371099	GU371163
Arum korolkowii	Boyce, RBG-Kew, 1994-3354	Tadzhikistan	GU371000	GU371037	GU371100	GU371164
Arum lucanum1	Boyce, RBG-Kew, 1987-1133	Italy	GU371001	GU371039	GU371102	GU371166
Arum lucanum2	Boyce, RBG-Kew, 1991-887	Italy	GU371002	GU371051	GU371114	GU371176
Arum maculatum1	Chase, RBG-Kew, 11161K	England	GU371003	GU371040	GU371103	GU371167
Arum maculatum2	NA, RBG-Kew, MJC 0002	England	GU371004	GU371041	GU371104	GU371168
Arum maculatum3	Boyce, RBG-Kew, 1990-475	Wales	GU371005	GU371042	GU371105	GU371169
Arum maculatum4	Boyce, RBG-Kew, 1990-2018	Turkey	GU371006	GU371043	GU371106	GU371170
Arum maculatum5	Bedalov & Küpfer, pr. coll., 1914b	Crimea	GU371007	GU371066	GU371129	GU371190
Arum maculatum6	Espíndola & Zryd, pr. coll., macKIL3	England	GU371008	GU371087	GU371151	GU371211
Arum maculatum7	Espíndola & Zryd, pr. coll., macLAC2	Italy	GU371009	GU371088	GU371152	GU371212
Arum maculatum9	Espíndola & Zryd, pr. coll., mac SFG7	France	GU371011	GU371090	GU371154	GU371214
Arum megobrebi1	Bedalov, pr. coll., 1432	Georgia	GU371012	GU371058	GU371121	GU371183
Arum megobrebi2	Bedalov, pr. coll., 1436	Georgia	GU371013	GU371059	GU371122	GU371184
Arum megobrebi3	Bedalov, pr. coll., 1442	Georgia	GU371014	GU371071	GU371135	GU371196
Arum megobrebi4	Neumann, pr. coll., 24219	Georgia	GU371015	GU371075	GU371139	GU371199
Arum nigrum	Boyce, RBG-Kew, 1992-2083	Montenegro	GU371016	GU371044	GU371107	GU371171
Arum orientale ssp. *longispathum*	Bedalov & Küpfer, pr. coll., 13.05.02 98/15	NA	GU371018	GU371070	GU371134	GU371195
Arum orientale1	Boyce, RBG-Kew, NA	NA	GU371019	GU371045	GU371108	GU371172
Arum orientale2	Bedalov & Küpfer, pr. coll., 284 99/61	Macedonia	GU371017	GU371067	GU371130	GU371191
Arum palaestinum	Chase, RBG-Kew, 11016K	NA	GU371020	GU371046	GU371109	GU371173
Arum pictum	Chase, RBG-Kew, 11024K	Balearic	GU371021	GU371047	GU371110	–
Arum purpureospathum	Chase, RBG-Kew, 11023K	Crete	GU371022	GU371048	GU371111	–
Arum rupicola ssp. *rupicola*	Chase, RBG-Kew, 11036K	Jordan	GU371023	GU371052	GU371115	GU371177
Arum rupicola ssp. *virensens*	Chase, RBG-Kew, 11034K	Turkey	GU371024	GU371050	GU371113	GU371175
Arum sintenisii	NA, BG-Basle, NA	NA	GU371025	GU371062	GU371125	GU371187
Arum sp. nov.	Neumann, pr. coll., 26940	Russia	GU371027	GU371076	GU371140	GU371200
Biarum davisii	NA, RBG-Kew, 2002-2839	NA	GU370971	GU371057	GU371120	GU371182
Biarum dispar	NA, RBG-Kew, 1991-123	Italy	GU370972	GU371056	GU371119	GU371181
Dracunculus canariensis	NA, RBG-Kew, 1982-2008	NA	GU370987	GU371053	GU371116	GU371178
Dracunculus vulgaris1	NA, RBG-Kew, 86 – 3894	NA	–	GU371054	GU371117	GU371179
Dracunculus vulgaris2	NA, RBG-Kew, NA	England	GU370988	GU371055	GU371118	GU371180

Chapter two

Molecular relationships in the sub-family Psychodinae (Diptera: Psychodidae)

Anahí Espíndola, Sven Buerki, Anouchka Jacquier,
Jan Ježek and Nadir Alvarez

Molecular relationships in the sub-family Psychodinae (Diptera: Psychodidae)

Anahí Espíndola[1], Sven Buerki[2], Anouchka Jacquier[1], Jan Ježek[3] and Nadir Alvarez[4]

[1]Laboratory of Evolutionary Entomology, Institute of Biology, University of Neuchâtel. Emile-Argand 11, 2000 Neuchâtel, Switzerland.

[2]Molecular Systematics Section, Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, United Kingdom.

[3]Department of Entomology, National Museum of Prague (Natural History), Praha 4, Czech Republic.

[4] Department of Ecology and Evolution, Biophore Building, University of Lausanne, 1015 Lausanne, Switzerland.

Abstract

Taxonomic and phylogenetic relationships in the order Diptera are still unclear not only at lower but also at higher taxonomic levels, as for instance in the family Psychodidae (lower Diptera). This family is formed by six subfamilies, of which only one (Phlebotominae) has been previously investigated on phylogenetic grounds. Among the remaining groups, the subfamily Psychodinae has only recently interested entomologists, and despite its worldwide distribution, little is known about their taxonomic and evolutionary relationships.

In the present study, we investigated the phylogeny of the subfamily using two molecular markers, by focusing on European/Palearctic species. A phylogenetic hypothesis as well as a temporal framework are inferred through the use of a dating approach that accounts for phylogenetic uncertainty.

Our results demonstrate that only one currently defined tribe (Psychodinii) is supported by molecular data as soon as it is merged with subtribe Brunettiina. Though tribes are generally not well supported, this is not the case of sub-tribes *sensu* Ježek, which were partly retrieved by our analyses. New fossil evidence allowed estimating the origin of the subfamily to the Late Cretaceous followed by the emergence of major clades and most current species during the Paleogene. Unexpectedly, a long morphological stasis is observed since all species diverged before the mid-Neogene, with the exception of two pairs of paraphyletic species. The impact of the Cretaceous-Paleogene (K-T) massive extinction on the diversification of the subfamily is also discussed.

Thanks to the application of this molecular approach to the investigation of the relationships in the subfamily Psychodinae, we finally propose some future lines of research to establish a new formal classification at the tribal and subtribal levels.

Introduction

The insect order Diptera (true flies) includes over 150'000 species, around 180 families and is considered one of the largest and most diverse groups of organisms (Bertone and Wiegmann 2009). In addition of being highly diversified, this order has also encountered an incredible success by colonizing all ecosystems (Yeates and Wiegmann 1999). Such ability was probably facilitated by the fact that species within this order are adapted to a wide diversity of ecological niches and trophic levels (*i.e.*, predators, decomposers, parasitoids and pollinators) (Bertone and Wiegmann 2009).

Despite their extreme abundance and species richness (several authors advocate that the species diversity is twice the number presented above), the systematics of Diptera remains problematic and therefore is only partially established. During the last century, taxonomists (e.g., Hennig 1976; Hennig 1981) have focused their research on publishing treatments for families harboring high species richness and/or showing some economic interest. Indeed, a non negligible species number of Diptera are vectors of human and animal pathogens (*e.g.*, Culicidae, Psychodidae) and many are pests of crops (*e.g.* Anthomyiidae, Tephritidae) (Bertone and Wiegmann 2009). Traditionally, classifications identified characters related to the antennae of adults and the head capsule of larvae and used them to subdivide the order into two suborders: Nematocera ("thread-horn" flies) and Brachycera ("short-horn" flies) (Yeates and Wiegmann 1999). More recently, the popularization of molecular methods has triggered the inference of phylogenetic frameworks for this group, which allowed testing previous classifications and using phylogenies as guidelines to establish new systematics when necessary (e.g., Friedrich and Tautz 1997; Bertone, Courtney et al. 2008). However, the fact that *ca.* 180 families were described within Diptera and that most of the families are poorly known (in terms of number of species, distributions, etc) have largely prevented the inference of accurate phylogenetic hypotheses. Today, while the monophyly of Brachycera is supported by molecular and morphological evidence, this suborder is considered to have originated from Nematocera, currently regarded as a set of paraphyletic infraorders (e.g., Bertone, Courtney et al. 2008). In addition

of being paraphyletic, evolutionary relationships with the lower Diptera (that partially corresponds to previous Nematocera) are not resolved and disagree with previous classifications. Finally, as reviewed by Bertone and Wiegmann (2009), fossil and molecular data support an origin of the order Diptera sometime during the Late Paleozoic [270-251 million years ago (Mya)] with divergence among families taking place in the Triassic and Jurassic (251-146 Mya).

To provide a phylogenetic insight into the relationships of some of the families, we focus our analyses on the evolutionary associations within the worldwide spread family Psychodidae (lower Diptera), comprising currently *ca.* 3000 species (Ježek and Barták 2000). This family is divided into six subfamilies (Bruchomyiinae, Trichopmyiinae, Horaiellinae, Phlebotominae, Psychodinae and Sycoracinae), with only some of them present in Europe. Current molecular clocks estimate the origin of the family to date back to the Cretaceous (ca. 140 Mya, Bertone and Wiegmann 2009), whereas fossil evidence [not taken into account for the dating of Diptera mentioned in Blagoderov *et al.* (2007)] suggests an earlier origin during the Late Triassic-Early Jurassic (*ca.* 201 Ma). Within the family, a phylogenetic hypothesis was only inferred for the subfamily Phlebotominae, a well-known taxon because of the ability of several species to carry human diseases (Depaquit, Perrotey et al. 1998; Beati, Caceres et al. 2004). In order to contribute to the knowledge of the evolution of the family, we focus in this survey on another subfamily: the Psychodinae, and most precisely on the study of samples inhabiting the European/Palearctic portion of its distribution.

Subfamily Psychodinae, presents a broad morphological and ecological diversity that allowed species from this group to disperse worldwide and to colonize and survive in remote places such as oceanic islands and subantarctic regions (Withers 1988). Since the last century, Vaillant (1971) and Ježek (1984) have provided a strong species-level taxonomic framework. For instance, Vaillant (1971) attempted to describe and normalize characters for classification of larvae and proposed identification keys to relate them to their respective imagi, while Ježek (1984) reorganized nomenclatorial problems and described species for

regions little investigated. Despite these works, all supra-specific entities have been subject to controversy and even challenged the first treatment done by Enderlein (1936) (several lumpings and splittings of taxonomic entities have been proposed; see Table 1). First, Vaillant (1971) recommended to subdivide the Palearctic Psychodinae into four tribes: Telmatoscopini, Psychodini, Pericomini and Brunettiini. Unfortunately, most of Vaillant's work lacked of nomenclatural (several types were wrongly cited) and cladistic rigor (relationships between species and genera in trees were not clarified), and Ježek (1983; 1990) thus proposed an alternative treatment based on a cladistic approach.

Although the circumscription of suprageneric levels of classification within the subfamily remains a challenging task, the strong taxonomic background available in Palearctic species (as well as the cladistic hypotheses provided by Ježek, 1983; 1990) allow confronting the past view of the Psychodinae systematics with a molecular approach. Based on a sampling covering all European tribes of Psychodinae, we use molecular sequencing of mitochondrial regions to infer phylogenetic relationships within the subfamily. The aims of this study are i) to evaluate the validity of Palaearctic tribes *sensu* Ježek or Vaillant; ii) when possible, to test the monophyly of genera; iii) to estimate the temporal origin of the subfamily and of the main tribes; iv) to estimate the age of current species in the context of a group that spanned the last 250 Mya; v) to propose guidelines for future classification within Psychodinae.

Material and Methods

Sampling

The sampling comprising 52 species [distributed into 26 genera, representing all tribes of the European Psychodinae (Ježek 2009) and including 87 samples] (see Supplementary Material 1 for information on success of amplification of samples) was obtained from the personal collection of Dr. J. Ježek. Samples were collected during several field trips, from years 1995 to 2009 in Eastern Europe, and were preserved in 70% ethanol. Seven outgroup species were added to the dataset based on recent phylogenetic studies on Diptera (see e.g., Bertone, Courtney et al. 2008). The outgroup species are distributed as follows: six

species belonging to the infraorder Culicomorpha (*Anopheles funestus, A. gambiae, A. quadrimaculatus, A. darling, Aedes albopictus* and *A. aegypti*) and one species belonging to the subfamily Phlebotominae (*Phlebotomus riouxi*).

DNA extraction, amplification and sequencing

DNA from insects was extracted using the QIAgen DNEasy Animal tissue extraction kit (QIAGEN, Hombrechtikon, Switzerland). Two mitochondrial regions (hereafter mtDNA), NADH1 and Cytochrome B (Cytb), were amplified with primers described in Simon *et al.* (1994) (N1-J-12248 -AAG CTA ATC TAA CTT CAT AAG- and LR-N-12866 -ACA TGA TCT GAG TTC AAA CCG G- for NADH1, and CB-J-11338 -CAC ATT CAA CCA GAA TGA TAT TT- and N1-N-12051 -GAT TTT GCT GAA GGT GAA TCA GA- for Cytb). Amplifications were performed in a 20µl master mix consisting of 0.5X buffer, 100mM dNTP, 0.5 µM of each primer, 0.12mM $MgCl_2$,1 unity of GoTaq Taq-polymerase (Promega, Switzerland) and 4 µl DNA. Reactions were run in a Biometra Thermocycler (Biometra, Goettingen, Germany). An initial denaturation of 1:30 min at 95°C, followed by 35 cycles of 1 min of denaturation at 95°C, 1 min of annealing at temperatures comprised between 42°C and 52°C, depending of samples and primers, and 30 s of elongation at 72°C. Cycles were followed by 5 min of final elongation at 72°C. The purification of PCR products and fluorescence sequencing were done by Macrogen Inc. (South Korea) and Fasteris SA (Switzerland) with the same primers as used for the PCR amplification.

Sequence alignment and phylogenetic reconstruction

The program ChromasPro 1.41 (Technelysium Pty Ltd.) was used to assemble complementary strands and verify software base-calling. The two mtDNA regions were initially aligned applying the Clustal-Wallis algorithm (implemented in BioEdit 7.0.4.1; Hall 1999) and further visually adjusted. For each mtDNA region gaps were coded in FastGap 1.2 (Borchsenius 2009), applying the simple gap-coding algorithm of Simmons and Ochoterena (2000). Finally, the number of constant, variable, parsimony informative and

151 total base pairs were calculated for each partition using MEGA 4.0.1 (Tamura, Dudley et al.
152 2007).

153 Before computing total evidence analyses, an incongruence length difference (ILD) test
154 (Farris, Källersjö et al. 1994) was performed using PAUP* 4b10 (Swofford 2003) with 100
155 replicates. If the probability value provided by the analysis is higher than 0.05, it is
156 statistically correct to assume that the information provided by both partitions is congruent
157 (Swofford 2003). Because the ILD test strongly rejected the incongruence between the
158 datasets (*p*-value=1), both regions were combined and total evidence analyses were
159 performed.

160 Phylogenetic relationships were inferred based on both probabilistic [*i.e.*, Bayesian
161 inference and Maximum Likelihood (ML)] and cladistic [*i.e.*, Maximum Parsimony (MP)]
162 approaches. A partitioned Bayesian analysis was performed using a Markov-Chain Monte-
163 Carlo (MCMC) approach, as implemented in MrBayes v3.1.2 (Ronquist and Huelsenbeck
164 2003) and following Nylander *et al.* (2004). Best-fit models for each region were selected
165 using MrModeltest v.3.0 (Nylander 2004) based on the Akaike information criterion (Akaike
166 1973). For both mtDNA partitions, the best-fit model was the general time reversible (GTR)
167 model with an alpha parameter for the shape of the gamma distribution to account for rate
168 heterogeneity among sites (Yang 1993). Since datasets were highly congruent (see above)
169 gaps from both mtDNA regions were merged into a single partition and the "restriction model"
170 was applied. Two Metropolis-coupled Markov chains with incremental heating temperature of
171 0.5 were run for 15 million generations and sampled every 1000th generation. The analysis was
172 repeated twice, starting from random trees. Convergence was accepted when standard
173 deviations attained values below 0.01 and when the Potential Scale Reduction Factor index
174 (Gelman and Rubin 1992) approached 1.0. We considered the MCMC sampling sufficient
175 when the Effective Sampling Size was higher than 200 –checked on Tracerv1.4 (Rambaut
176 and Drummond 2004). After a burn-in period corresponding to 5'000 trees a half-compatible
177 consensus tree (and its associated Bayesian posterior probabilities; BPP) was reconstructed

178 (based on the remaining 10’000 trees) using MrBayes v3.1.2 (Ronquist and Huelsenbeck
179 2003).
180 The ML analysis was done using RAxML 7.2.6 (Stamatakis 2006) with 10’000 rapid
181 bootstrap analyses followed by the search of the best-scoring ML tree in one single run. The
182 two mtDNA regions were considered as one single partition (by applying the GTR model)
183 and the gap partition was not considered. The analysis was done using the facilities offered
184 by the CIPRES portal (San Diego, CA, USA).
185 The combined dataset was analyzed using parsimony ratchet (Nixon, 1999) with PAUPrat
186 (Sikes and Lewis 2001). Based on recommendations of Nixon (1999), ten independent
187 searches (with 200 iterations and 15% of the parsimony informative characters perturbed)
188 were conducted. The most parsimonious trees were afterwards combined to create a
189 majority-rule consensus tree. Branch supports were estimated using the Decay Index
190 (Bremer 1988) as implemented in TreeRot v3 (Sorenson and Franzosa 2007). The Decay
191 Index indicates the number of steps necessary to collapse a node, informing on the stability
192 and consistency of each clade (Felsenstein 2004). The congruence of topological groupings
193 between the Bayesian, ML and MP algorithms was evaluated by visually comparing
194 topologies and node supports.
195
196 *Estimation of lineage divergence times*
197 Relative branching times were estimated on the halfcompat consensus tree of MrBayes
198 using Penalized Likelihood (PL; Sanderson 2002), as implemented in r8s v.1.71 (Sanderson
199 2004), and the Truncated Newton method algorithm. The smoothing value (10) was
200 established using the cross-validation routines implemented in r8s. One of the most external
201 outgroups, *Aedes aegypti*, was pruned for the estimation of the divergence time as required
202 by the program (see Sanderson 2004).
203 To account for phylogenetic uncertainty in the estimation of lineage divergence times, PL
204 analyses were performed on a random sample of trees (N = 4000) from the Bayesian MCMC
205 stationary distribution as done in Buerki et al. (in press). This was done as follows: i) PL

206 analyses were run independently on the set of trees using the same smoothing parameter
207 value as the consensus tree; ii) 95% confidence intervals of age estimates for each node
208 compatible with the halfcompat tree were calculated based on the sample of PL-dated trees;
209 iii) the intervals of confidence were finally represented on the consensus tree. These steps
210 were done using a set of R scripts (R Development Core team, 2009) available on request to
211 SB.

212 To estimate absolute ages for lineage divergences, three fossil calibration points (see
213 below) were used to set minimum age constraints for several nodes in the phylogeny (the
214 root was estimated using a maximum node age; see below). The calibration procedure was
215 done as in Buerki et al. (in press). For each calibration point, the oldest fossil record was
216 selected and the upper (younger) bound of the geological interval (Walker and Geissman
217 2009) in which the fossil was found was used to represent the minimum age constraint.

218 Calibration points were defined as follows: (a) the root node – *i.e.*, the most recent
219 common ancestor (MRCA) of Diptera – was constrained to a maximum age of 260 million
220 years (Mya). According to Bertone *et al.* (2008) fossil evidences supporting this calibration
221 point were estimated from the Capitanian to the Wuchiapingian (Mid- Late Permian). (b) The
222 stem group of the Psychodidae was set to a minimum age of 201.6 Ma. This calibration point
223 is based on fossil evidence (Blagoderov, Grimaldi et al. 2007) occurring from the Rhaetian to
224 the Hetangian (Late Triassic-Early Jurassic). (c) Finally, the stem group of Psychodinae was
225 set to a minimum age of 176 Ma. This point is based on fossils (Ansorge 1994) found for the
226 Toarcian to the Aalenian periods (Early-Mid Jurassic).

227 Finally, in order to evaluate the consistency and age of species, their monophyly was
228 retrieved from the Bayesian topology and the time between the species' MRCA and the
229 anterior split of the lineage (*i.e.*, the age of the node anterior to the MRCA, corresponding to
230 the split with the sister species) was calculated and recorded for each species using scripts
231 available to SB. Classes of five million years were considered to reconstruct histograms for
232 the visualization of the distribution of species divergence through time.

233

234 **Results**

235 Among 94 sequenced samples (87 in the ingroup and seven outgroups), 83 properly
236 amplified region Cytb and 57 region NDHA1 (Supplementary Material 1). The combined
237 aligned dataset was 1248bp length: 652bp for Cytb and 596bp for NADH1. A gap partition of
238 87 characters (coded as presence/absence) was added to the matrix. The number of
239 variable, constant and potentially parsimony-informative sites are summarized in Table 2.
240 The number of variable sites were highly similar between the two mtDNA regions and almost
241 50% of sites were parsimony-informative (Table 2).

242 The dating analysis (performed on 4000 trees randomly selected from the Bayesian
243 posterior probability; Figure 1) estimated the origin of the MRCA of Psychodinae as having
244 happened during the Late Cretaceous (*ca.* 85.6 Mya), and the stem of the subfamily going
245 back to the Mid Jurassic (*ca.* 176 Mya) (Figure 1). Within Psychodinae, lineages diverged
246 between the Late Cretaceous (*ca.* 75.01 Mya) and the Early Paleogene (59.97 Ma), with
247 most of the differentiation taking place during the Paleogene. The ages of species' lineages
248 were unexpectedly ancient, with most stems originating in the Paleogene, and no lineages
249 more recent than 10 Ma (Figure 3).

250 The three phylogenetic approaches provided partially congruent topologies (Figure 2 and
251 Supplementary Material 2). We defined main clades at higher taxonomic levels, by using a
252 temporal cut-off at 75 Ma. Four clades were thus retrieved, namely I to IV (Figure 2). While
253 clades I, II and III were also retrieved in the ML and MP approaches, this was not the case of
254 clade IV, which was paraphyletic with regard to the other clades in the MP approach.
255 Because clades II and IV comprised highly diverging lineages, we categorized taxa one step
256 further and considered five subclades in each of these two clades (respectively IIa to IIe, and
257 IVa to IVe) (Figure 2). These subclades were defined on a stem age ranging between 46
258 and 60 Ma. While half of them were monophyletic whatever approach used, the other half
259 showed a looser level of congruence among phylogenetic approaches; node supports varied
260 as well, ranging from low to high values (see Table 3 and Figure 2). Monophyly irrespective
261 of the phylogenetic criterion was considered as strong evidence to validate a given clade. At

the generic level, seven of the twelve genera for which more than one species was sampled were monophyletic in at least one criterion of phylogenetic reconstruction (Table 4). The remaining five genera were para- or polyphyletic whatever criterion applied.

Discussion

Intricate relationships between taxonomy and phylogeny

Results presented in this study i) support the monophyly of subfamily Psychodinae, ii) indicate the polyphyly of all tribes (with the only exception of Psychodini that is paraphyletic with subtribe Brunettiina; see below) and iii) challenge the monophyly of several genera (Figures 1 and 2).

To avoid misleading relationships caused by putative high saturations of the mtDNA regions (see below), we discuss in this study only genera para/polyphyletic under the three phylogenetic algorithms. From this perspective, five genera are either polyphyletic (*Logima*) or paraphyletic (*Pericoma*, *Pneumia*, *Psychoda* and *Ulomyia*) (Table 4). The polyphyletic genus *Logima* is retrieved in two clades (IId and IIe), whereas the paraphyletic genera are restricted to one clade (IVe). In the latter case, further investigations are required before concluding on the status of the genera.

Although not in perfect agreement with the phylogenetic relationships, two of the four subtribes defined by Ježek match our phylogenetic clades: Brunettiina (clade IIb) and Mormiina (clade IVc) (Table 3). The two remaining subtribes, Paramormiina (clade IV) and Trichopyschodina (clades I and III) are paraphyletic (Figure 2 and Table 3). These new findings suggest that part of the taxonomy/systematics of Ježek (based on cladistic analyses) tends to reflect the evolutionary history of Psychodinae and should be considered when proposing a new formal classification. In addition, even if Trichopsychodina is paraphyletic, the two lineages including representatives of this tribe comprise only genera *Threticus* (clade I) and *Trichopsychoda* (clade III) respectively, what might therefore also constitute valuable information to circumscribe future subtribes.

289 Despite genera of tribe Psychodini were not clearly assigned to any subclade, they were
290 all restricted to clade II. Considering this, our result constitutes a step towards the
291 establishment of a new classification that incorporates both morphological and molecular
292 evidence. Tribe Psychodini is however paraphyletic due to the inclusion of Brunettiina
293 (*Atrichobrunettia*), a subtribe that previously challenged taxonomists (Figure 2, Table 3).
294 Even though this subtribe has been recognized as a separate entity, but at different
295 taxonomic ranks according to the researcher (Vaillant 1971; Ježek 1984), our analysis
296 supports its relationships with members of tribe Psychodini (Figures 1 and 2; Table 3).
297 Further investigations will be required to identify morphological synapomorphies supporting
298 the definition of these subclades. Based on our results, it appears that subtribe Brunettiina
299 presents an evolutionary history extremely related to that of tribe Psychodini. We thus
300 recommend taxonomists to adopt a broad concept of tribe Psychodini by also including
301 subtribe Brunettiina, since this would reflect historical and evolutionary relationships between
302 the two groups.
303 Finally, the remaining clade IV (Figures 1 and 2) is the most problematic when compared
304 with current classifications. With the only exception of subtribe Mormiina, none of the
305 published taxonomic entities corresponds to these phylogenetic subclades (Table 3). Tribe
306 Pericomini is retrieved in two non-sister clades (IVb and IVe) suggesting the need of
307 providing a new circumscription for this group. Moreover, a similar situation is observed with
308 subtribe Paramormiina (clades IVa, IVd). Although clades are usually moderately to strongly
309 supported, phylogenetic relationships within clade IV are not clear and would deserve
310 additional investigations before allowing further taxonomic recommendations.
311 To solve phylogenetic relationships (*e.g.*, basal polytomy of clades I, II, III, IV) and
312 recognize morphological synapomorphies supporting clades (at the basis of a new formal
313 classification) we suggest that researchers should i) increase their sampling to balance the
314 number of representatives at the different taxonomic levels, ii) sequence less polymorphic
315 markers (*e.g.*, 28S) and iii) compile an extensive morphological matrix. Although paradoxical
316 at first thought, the sequencing of less polymorphic regions will allow solving basal

polytomies that are most likely caused by high levels of saturation of the molecular regions used in this study (data not shown). This putative high level of saturation might also be the reason of the incongruence found between algorithms, especially between probabilistic algorithms and MP inferences (*e.g.* Clade IV). Probabilistic algorithms -through the use of evolutionary models- appear thus to be better suited to handle such situations and should be preferred over the MP approach in the case of this group.

The timing of diversification in Psychodinae

Although our dating inference might be biased by taxon sampling, the new fossil evidence (Blagoderov, Grimaldi et al. 2007) used here (compared to Bertone and Wiegmann 2009) allows estimating the origin of subfamily Psychodinae to sometime during the Late Cretaceous (*ca.* 85 Mya; Figure 1), with an emergence of all major clades during the Paleogene (65.5-23 Mya; Figures 1 and 3). Our results also support a diversification of all monophyletic genera during the Eocene (*ca.* 35 Mya) and, more strikingly, the emergence of sampled species during the Miocene (before 10 Mya; Fig. 3). Once all lineages diversified, a long period of "morphological" stasis is observed.

From a historical point of view, our results show that clades appear to have diverged during the Paleogene, a period directly following the now well-known K-T mass-extinction (around 65 Mya). This mass-extinction event was extraordinary and has been well identified in the fossil register, where the extinction of high proportions (*i.e.* up to 95% in some taxa) of species was observed in short geological times (Jablonski 2001). The period following this massive extinction has been punctuated in several groups by a rapid diversification, explained generally by the fact that massive extinctions left "free" environments with changed ecological networks, in which species found ample niches to evolve in (Erwin 2001). It is thus interesting to note that the origin of tribe Psychodini *sensu lato* (including Brunettiina) coincides with this post-K-T period (with a MRCA at 64 Mya). Moreover, almost all main groups in the subfamily appear to have diversified during the period directly following the *pulse* K-T extinction (sensu Erwin 1996), which would be in agreement with the

idea of post-massive extinction radiations with differential lag-times depending on the organism studied (Erwin 2001).

The ancient origin of Psychodinae, together with the observed “morphological” stasis are additional clues supporting the intricate evolutionary history of this group. These evidences (probably combined with high extinction events as those proposed above) appear thus as additional explanations for the multiple obstacles that have up to now prevented the development of a convincing classification for the group. In this context, our study constitutes a leap forward in the understanding of the evolution of Psychodinae and provides useful clues for a new formal classification. We are convinced that the circumscription of high taxonomic ranks within the subfamily should also be based on dating and evolutionary grounds, especially when considering the difficulties experienced when trying to define taxonomic borders using morphological characters only.

Acknowledgments

Authors would like to thank the technical help of Laure Benoit (CNRS Montpellier) for the sequencing of some samples. This study was funded by the Swiss National Science Foundation (project No. 3100A0-116778).

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. Second International Symposium on Information Theory. A. Kiado. Budapest**:** 267-281.

Ansorge, J. (1994). "Tanyderidae and Psychodidae (Insecta: Diptera) from Lower Jurassic of Northeastern Germany." Paläontologische Zeitschrift **68**: 199-210.

Beati, L., A. G. Caceres, et al. (2004). "Systematic relationships among *Lutzomyia* sand flies (Diptera: Psychodidae) of Peru and Colombia based on the analysis of 12S and 28S ribosomal DNA sequences." International Journal of Parasitology **34**(2): 225-34.

Bertone, M. A., G. W. Courtney, et al. (2008). "Phylogenetics and temporal diversification of the earliest true flies (Insecta: Diptera) based on multiple nuclear genes." Systematic Entomology **33**: 668-687.

Bertone, M. A. and B. M. Wiegmann (2009). True flies (Diptera). The Timetree of Life. S. B. Hedges and S. Kumar. UK, Oxford University Press**:** 270-277.

Blagoderov, V., D. A. Grimaldi, et al. (2007). "How Time Flies for Flies: Diverse Diptera from the Triassic of Virginia and Early Radiation of the Order." American Museum Novitates **3572**: 1-39.

Borchsenius, F. (2009). FastGap 1.2. Department of Biological Sciences, University of Aarhus, Denmark, Published online at http://192.38.46.42/aubot/fb/FastGap_home.htm.

Bremer, K. (1988). "The limits of amino acid sequence data in angiosperm phylogenetic reconstruction." Evolution **42**(795-803): 795.

Buerki, S., F. Forest, et al. (in press). "An evaluation of new parsimony-based versus parametric inference methods in biogeography: a case study using the globally distributed plant family Sapindaceae." Journal of Biogeography.

Depaquit, J., S. Perrotey, et al. (1998). "Molecular systematics of Phlebotominae: a pilot study. Paraphyly of the genus *Phlebotomus*." Comptes Rendus de l'Académie des Sciences III **321**(10): 849-55.

Erwin, D. H. (1996). Understanding biotic recoveries: extinction, survival, and preservation during the End-Permian mass extinction. Evolutionary paleobiology. D. Jablonski, D. H. Erwin and L. Lipps. Chicago, University Chicago Press 398-418.

Erwin, D. H. (2001). "Lessons from the past: biotic recoveries from mass extinctions." Proceedings of the National Academy of Sciences of the United States of America **98**(10): 5399-403.

Farris, J. S., M. Källersjö, et al. (1994). "Testing significance of incongruence." Cladistics(10): 315-319.

Fauna Europaea Web Service (2004). Fauna Europaea version 1.1.

Felsenstein, J. (2004). Inferring Phylogenies. Sunderland, MA, Sinauer Associates.

Friedrich, M. and D. Tautz (1997). "Evolution and phylogeny of the Diptera: a molecular phylogenetic analysis using 28S rDNA sequences." Systematic Biology **46**(4): 674-98.

Gelman, A. and D. B. Rubin (1992). "Inference from iterative simulation using multiple sequences." Statistical Science **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 Symposium Series **41**: 95-98.

Hennig, W., Ed. (1976). Anthomyiidae. Die Fliegen der Palaearktischen Region. Stuttgart, E. Schweizerbart.

Hennig, W. (1981). Insect Phylogeny. New York, Wiley.

411 Jablonski, D. (2001). "Lessons from the past: evolutionary impacts of mass extinctions."
412 Proceedings of the National Academy of Sciences of the United States of America
413 **98**(10): 5393-8.
414 Ježek, J. (1983). "Intergeneric relationships of selected tribes of the subfamily Psychodinae
415 (Diptera, Psychodidae)." Acta Entomologica Musei Nationalis Pragae **41**: 255-259.
416 Ježek, J. (1984). "Nomenclatorical changes of some higher taxa of palaearctic Psychodinae
417 (Diptera, Psychodidae)." Acta Faunistica Entomologica Musei Nationalis Pragae
418 **17**(199): 155-170.
419 Ježek, J. (1990). "Contribution to the taxonomy of some genera of Paramormiine moth flies
420 (Diptera, Psychodidae) with description of a new genus *Karakovounimerus*." Acta
421 Entomologica Musei Nationalis Pragae **43**: 129-157.
422 Ježek, J. (2009). "Psychodidae Newman 1834." Checklist of Diptera of the Czech Republic
423 and Slovakia. Electronic version 2, from http://zoology.fns.uniba.sk/diptera2009.
424 Ježek, J. and M. Barták (2000). "Psychodidae." Folia Facultatis Scientiarium Naturalium
425 Universitatis Masarykianae Brunensis **104**: 93-100.
426 Nixon, K. C. (1999). "The parsimony ratchet, a new method for rapid parsimony analysis."
427 Cladistics **15**: 407-414.
428 Nylander, J. A. A. (2004). MrModeltest v2. Uppsala, Sweden, Program distributed by the
429 author.
430 Nylander, J. A. A., F. Ronquist, et al. (2004). "Bayesian phylogenetic analysis of combined
431 data." Systematic Biology **53**: 47-67.
432 Rambaut, A. and A. J. Drummond (2004). Tracer v1.4. Edinburgh.
433 R Development Core Team (2009). R: A Language and Environment for Statistical
434 Computing.
435 Ronquist, F. and J. P. Huelsenbeck (2003). "MrBayes 3: Bayesian phylogenetic inference
436 under mixed models." Bioinformatics **19**(12): 1572-4.
437 Sanderson, M. J. (2002). "Estimating absolute rates of molecular evolution and divergence
438 times: a penalized likelihood approach." Molecular Biology and Evolution **19**: 101-
439 109.
440 Sanderson, M. J. (2004). r8S, program and documentation.
441 Sikes, D. S. and P. O. Lewis (2001). PAUPRat, Beta software, version 1: PAUP
442 implementation of the parsimony ratchet, Distributed by the authors.
443 Simmons, M. P. and H. Ochoterena (2000). "Gaps as characters in sequence-based
444 phylogenetic analyses." Systematic Biology **49**(2): 369-81.
445 Simon, C., F. Frati, et al. (1994). "Evolution, Weighting, and Phylogenetic Utility of
446 Mitochondrial Gene-Sequences and a Compilation of Conserved Polymerase Chain-
447 Reaction Primers." Annals of the Entomological Society of America **87**(6): 651-701.

448 Sorenson, M. D. and E. A. Franzosa (2007). Treerot, version 3. Boston MA, USA, Distributed
449 by the authors.
450 Stamatakis, A. (2006). "RAxML-VI-HPC: Maximum Likelihood-based Phylogenetic Analyses
451 with Thousands of Taxa and Mixed Models." Bioinformatics **22**(21): 2688-2690.
452 Swofford, D. L. (2003). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other
453 Methods). Version 4. Sunderland, Massachusetts, Sinauer Associates.
454 Tamura, K., J. Dudley, et al. (2007). "MEGA4: Molecular Evolutionary Genetics Analysis
455 (MEGA) software version 4.0." Molecular Biology and Evolution **24**: 1596-1599.
456 Vaillant, F. (1971). 9d. Psychodidae-Psychodinae. Die Fliegen der Palaearktischen Region.
457 E. Lindner. Stuttgart, E. Schweizerbartische Verlagsbuchhandlung. **9d:** 358.
458 Walker, J. D. and J. W. Geissman (2009). Geologic Time Scale: Geological Society of
459 America, The Geological Society of America.
460 Withers, P. (1988). "Moth Flies. Diptera: Psychodidae." Dipterist Digest **4**: 1-83.
461 Yang, Z. H. (1993). "Maximum-likelihood-estimation of phylogeny from DNA-sequences
462 when substitution rates differ over sites." Molecular Biology and Evolution **10**: 1396-
463 1401.
464 Yeates, D. K. and B. M. Wiegmann (1999). "Congruence and controversy: toward a higher-
465 level phylogeny of Diptera." Annual Review of Entomology **44**: 397-428.
466
467
468

469 Table 1 – Systematic position of all sampled taxa. Names are given following the taxonomy
470 proposed by Ježek (1984) and FAUNA EUROPAEA (Fauna Europaea Web Service, 2004). Supra-
471 generic classifiation is given following Ježek (1984) and Vaillant (1971). Phylogenetic classification
472 considering this work is given in the last column.

Species name (Ježek)	Species name (FAUNA EUROPAEA)	Ježek (1984)			Vaillant (1971)	This study
		Sub-family	Tribe	Sub-tribe	Tribe	
Threticus incurvus	*Threticus incurvus* Krek 1972	Psychodinae	Paramormiini	Trichopsychodina*	Telmatoscopini	I
Threticus lucifugus	*Threticus lucifugus* Walker 1856	Psychodinae	Paramormiini	Trichopsychodina*	Telmatoscopini	I
Threticus silvaticus	*Threticus silvaticus* Ježek 1986	Psychodinae	Paramormiini	Trichopsychodina*	Telmatoscopini	I
Psychodocha cinerea	*Psychoda cinerea* Banks 1894	Psychodinae	Psychodini*		Psychodini	IIa
Psychodocha gemina	*Psychoda gemina* Eaton 1904	Psychodinae	Psychodini*		Psychodini	IIa
Atrichobrunettia graeca	*Atrichobrunettia graeca* Ježek & Goutner 1993	Psychodinae	Mormiini	Brunettiina*	Brunettiini	IIb
Psychodula minuta	*Psychoda minuta* Banks 1895	Psychodinae	Psychodini*		Psychodini	IIc
Copropsychoda brevicornis	*Psychoda brevicornis* Tonnoir 1940	Psychodinae	Psychodini*		Psychodini	IId
Logima albipennis	*Psychoda albipennis* Zetterstedt 1852	Psychodinae	Psychodini*		Psychodini	IId
Logima satchelli	*Psychoda satchelli* Quate 1955	Psychodinae	Psychodini*		Psychodini	IId
Ypsydocha setigera	*Psychoda setigera* Tonnoir 1923	Psychodinae	Psychodini*		Psychodini	IId
Logima erminea	*Psychoda erminea* Eaton 1898	Psychodinae	Psychodini*		Psychodini	IIe
Psycha grisescens	*Psychoda grisescens* Tonnoir 1923	Psychodinae	Psychodini*		Psychodini	IIe
Psychoda crassipenis	*Psychoda crassipenis* Tonnoir *1940*	Psychodinae	Psychodini*		Psychodini	IIe
Psychoda phalaenoides	*Psychoda phalaenoides* Linnaeus 1758	Psychodinae	Psychodini*		Psychodini	IIe
Psychoda uniformata	*Psychoda uniformata* Haseman 1907	Psychodinae	Psychodini*		Psychodini	IIe
Psychomora mycophila	*Psychoda mycophila* Vaillant 1990	Psychodinae	Psychodini*		Psychodini	IIe
Tinearia alternata	*Tinearia alternata* Say 1824	Psychodinae	Psychodini*		Psychodini	IIe
Tinearia lativentris	*Tinearia lativentris* Berden 1952	Psychodinae	Psychodini*		Psychodini	IIe
Trichopsychoda hirtella	*Trichopsychoda hirtella* Tonnoir 1919	Psychodinae	Paramormiini	Trichopsychodina*	Telmatoscopini	III
Clogmia albipunctata	*Clogmia albipunctatus* Williston 1893	Psychodinae	Paramormiini	Paramormiina*	Telmatoscopini	IVa
Paramormia polyascoidea	*Paramormia polyscoidea* Krek 1970	Psychodinae	Paramormiini	Paramormiina*	Telmatoscopini	IVa
Paramormia ustulata	*Paramormia ustulata* Walker 1856	Psychodinae	Paramormiini	Paramormiina*	Telmatoscopini	IVa
Sciria advena	*Telmatoscopus advenus* Eaton 1893	Psychodinae	Paramormiini	Paramormiina*	Telmatoscopini	IVa
Clytocerus longicorniculatus	N/A	Psychodinae	Pericomini*		Pericomini	IVb
Clytocerus ocellaris	*Clytocerus) ocellaris* Meigen 1818	Psychodinae	Pericomini*		Pericomini	IVb
Clytocerus rivosus	*Clytocerus rivosus* Tonnoir 1919	Psychodinae	Pericomini*		Pericomini	IVb
Oomormia andrenipes	*Mormia andrenipes* Strobl 1910	Psychodinae	Mormiini	Mormiina*	Pericomini	IVc
Jungiella procera	*Jungiella procera* Krek 1971	Psychodinae	Paramormiini	Paramormiina*	Telmatoscopini	IVd
Jungiella valachia	*Jungiella valachia* Vaillant 1963	Psychodinae	Paramormiini	Paramormiina*	Telmatoscopini	IVd
Parajungiella consors	*Jungiella consors* Eaton 1893	Psychodinae	Paramormiini	Paramormiina*	Telmatoscopini	IVd
Panimerus denticulatus	*Panimerus denticulatus* Krek 1972	Psychodinae	Paramormiini		Telmatoscopini	IVd
Telmatoscopus britteni	*Telmatoscopus britteni* Tonnoir 1940	Psychodinae	Paramormiini	Paramormiina*	Telmatoscopini	IVd
Berdeniella illiesi	*Berdeniella illiesi* Wagner 1973	Psychodinae	Pericomini*		Pericomini	IVe
Berdeniella stavniensis	*Berdeniella stavniensis* Krek 1969	Psychodinae	Pericomini*		Pericomini	IVe
Berdeniella unispinosa	*Berdeniella unispinosa* Tonnoir 1919	Psychodinae	Pericomini*		Pericomini	IVe
Pericoma blandula	*Pericoma blandula* Eaton 1893	Psychodinae	Pericomini*		Pericomini	IVe
Pericoma exquisita	*Pericoma exquisita* Eaton 1893	Psychodinae	Pericomini*		Pericomini	IVe
Pericoma rivularis	*Pericoma rivularis* Berden 1955	Psychodinae	Pericomini*		Pericomini	IVe
Pneumia compta	*Satchelliella compta* Eaton 1893	Psychodinae	Pericomini*		Pericomini	IVe
Pneumia gracilis	*Satchelliella gracilis* Eaton 1893	Psychodinae	Pericomini*		Pericomini	IVe
Pneumia nubila	*Satchelliella mutua* Eaton 1893	Psychodinae	Pericomini*		Pericomini	IVe
Pneumia pilularia	*Satchelliella pilularia* Tonnoir 1940	Psychodinae	Pericomini*		Pericomini	IVe
Pneumia plumicornis	*Satchelliella plumicornis* Tonnoir 1922	Psychodinae	Pericomini*		Pericomini	IVe
Pneumia stammeri	*Satchelliella stammeri* Jung 1954	Psychodinae	Pericomini*		Pericomini	IVe
Pneumia trivialis	*Satchelliella trivialis* Eaton 1893	Psychodinae	Pericomini*		Pericomini	IVe
Saraiella rotunda	*Saraiella rotunda* Krek 1970	Psychodinae	Pericomini*		Pericomini	IVe
Ulomyia cognata	*Ulomyia cognata* Eaton 1893	Psychodinae	Pericomini*		Pericomini	IVe
Ulomyia fuliginosa	*Ulomyia fuliginosa* Meigen 1818	Psychodinae	Pericomini*		Pericomini	IVe
Ulomyia undulata	*Ulomyia undulata* Tonnoir 1919	Psychodinae	Pericomini*		Pericomini	IVe

473 NB. An "*" indicates the best-match subtribal or tribal classification following Ježek (1984).

Table 2 – Constant (C), variable (V), parsimony informative (PI) and total number of sites, in base pairs. Percentage of total base pairs are indicated in parenthesis. Gaps for each region are also shown, and total sequence length and number of gaps are indicated considering both regions together.

	C	V	PI	Total	Gaps
NADH-1	306 (51.34)	289 (48.49)	244 (40.94)	**596**	54
Cyt-b	264 (40.49)	388 (59.51)	332 (50.92)	**652**	33
Total				**1248**	**87**

Table 3 – Age of the Most Recent Common Ancestor (MRCA) of each group or subgroup identified as a monophyletic clade in the Bayesian inference analysis, with best-match taxonomic classification at the subtribal or tribal (between parentheses) levels, following Ježek (1984). When subtribes were revealed to be paraphyletic, a mention for each distinct group was added (§1 and §2). Bayesian Posterior Probability (BPP), Maximum Likelihood (ML) bootstraps and Maximum Parsimony (MP) Decay indexes are shown to inform on clade support. Groups or subgroups in bold indicate clades whose monophyly is supported by all phylogenetic criteria. Names of groups and subgroups correspond to those stated in Figures 1 and 2.

Group	Subgroup	Sub-tribal taxonomy according to Ježek (1984)	Age of stem of MRCA	BPP	ML support	MP Decay Index
I		**Trichopsychodina §1**	**85.6**	**0.99**	**73**	**12**
II			**85.6**	**0.99**	**67**	**11**
	IIa	*(Psychodini)*	*64.48*	*0.57*	*40*	*N/A*
	IIb	***Brunettiina***	***47.27***	***1***	***100***	***45***
	IIc	***(Psychodini)***	***47.27***	***0.85***	***84***	***18***
	IId	*(Psychodini)*	*54.11*	*0.93*	*26*	*N/A*
	IIe	*(Psychodini)*	*57.03*	*0.98*	*41*	*N/A*
III		**Trichopsychodina §2**	**75.01**	**1**	**100**	**0**
IV			75.01	0.96	47	N/A
	IVa	*Paramormiina §1*	*59.97*	*0.66*	*N/A*	*N/A*
	IVb	***(Pericomini)***	***59.97***	***0.99***	***87***	***6***
	IVc	***Mormiina***	***46.84***	***1***	***100***	**31**
	IVd	*Paramormiina §2*	*46.84*	*0.55*	*48*	*N/A*
	IVe	***(Pericomini)***	***53.19***	***0.93***	***29***	***1***

Table 4 – a) Summary of genera *sensu* Ježek (1984) for which monophyly could be retrieved with Bayesian inference (BI), Maximum Parsimony (MP) and Maximum Likelihood (ML) approaches. Number of species sampled per genus are also indicated. N/A : monophyly not tested because only one species sampled. b) Counts of monophyletic and paraphyletic genera.

a)

Genus	Number of species	Monophyly		
		Bayesian inference	Maximum Parsimony	Maximum Likelihood
Atrichobrunettia	1	N/A	N/A	N/A
Berdeniella	3	yes	yes	yes
Clogmia	1	N/A	N/A	N/A
Clytocerus	3	yes	yes	yes
Copropsychoda	1	N/A	N/A	N/A
Jungiella	2	yes	yes	yes
Logima	3	no	no	no
Oomormia	1	N/A	N/A	N/A
Panimers	1	N/A	N/A	N/A
Parajungiella	1	N/A	N/A	N/A
Paramormia	2	no	yes	no
Pericoma	3	no	no	no
Pneumia	8	no	no	no
Psycha	1	N/A	N/A	N/A
Psychoda	3	no	no	no
Psychodocha	2	yes	no	yes
Psychodula	1	N/A	N/A	N/A
Psychomora	1	N/A	N/A	N/A
Saraiella	1	N/A	N/A	N/A
Sciria	1	N/A	N/A	N/A
Telmatoscopus	1	N/A	N/A	N/A
Threticus	3	yes	yes	yes
Tinearia	2	yes	yes	yes
Trichopsychoda	1	N/A	N/A	N/A
Ulomyia	4	no	no	no
Ypsydocha	1	N/A	N/A	N/A

b)

Monophyly	Count
yes for BI, MP or ML (or all)	7 (5)
no for neither BI, MP and ML	5
N/A	14
Total	26

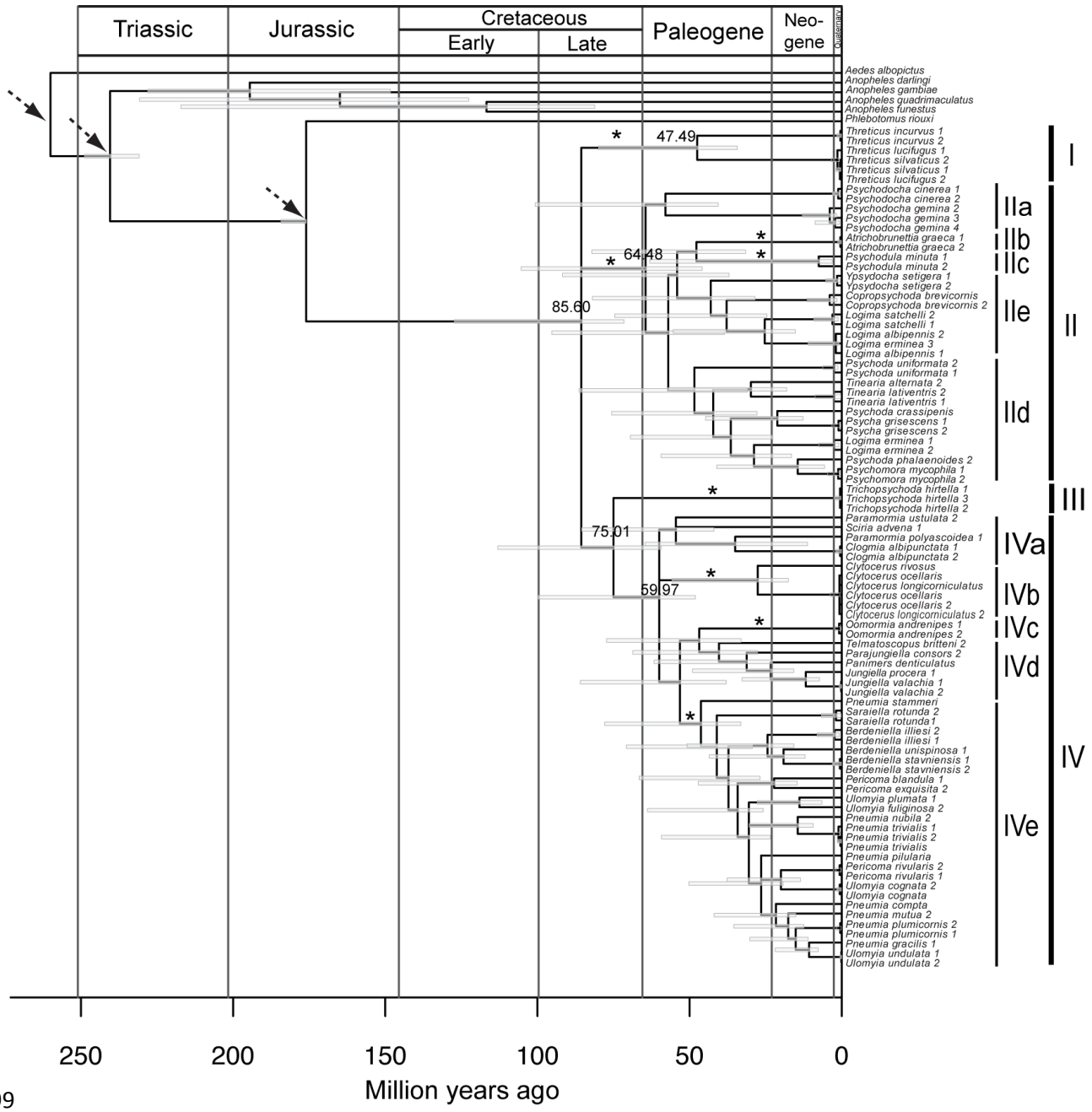


Figure 1 – Molecular dating of subfamily Psychodinae. Arrows indicate fossil calibration points. Dates on the topology indicate divergence times of the four main clades (I, II, III and IV). Vertical bars on the right illustrate the informal classification we propose to use here. Stars on a stem indicate a node that is supported by a monophyletic status whatever phylogenetic criterion is applied.

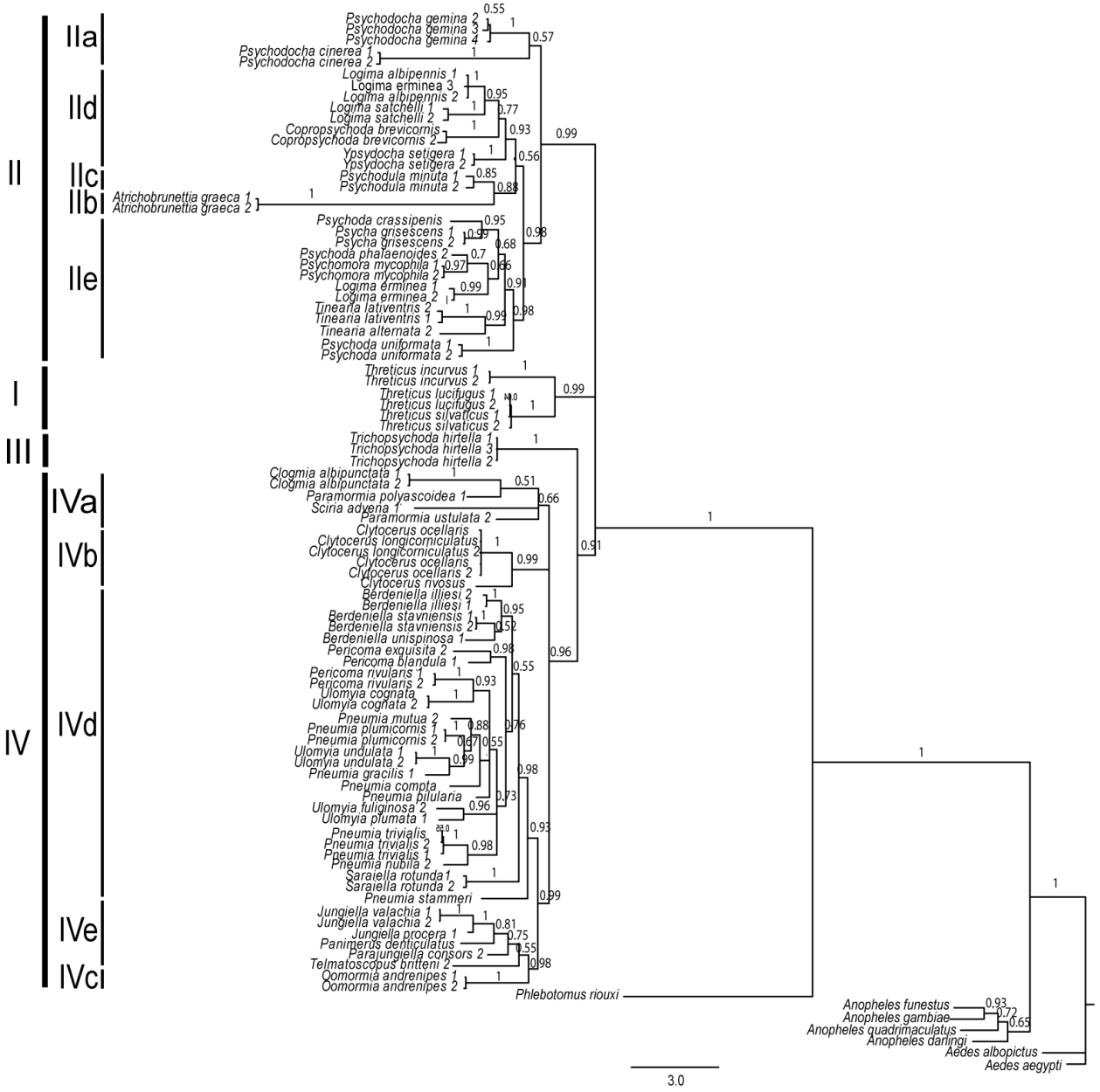


Figure 2 – Phylogeny of subfamily Psychodinae using Bayesian inference. Node supports refer to Bayesian Posterior Probability (BPP) values. Vertical bars on the left illustrate the classification considered in the present study.

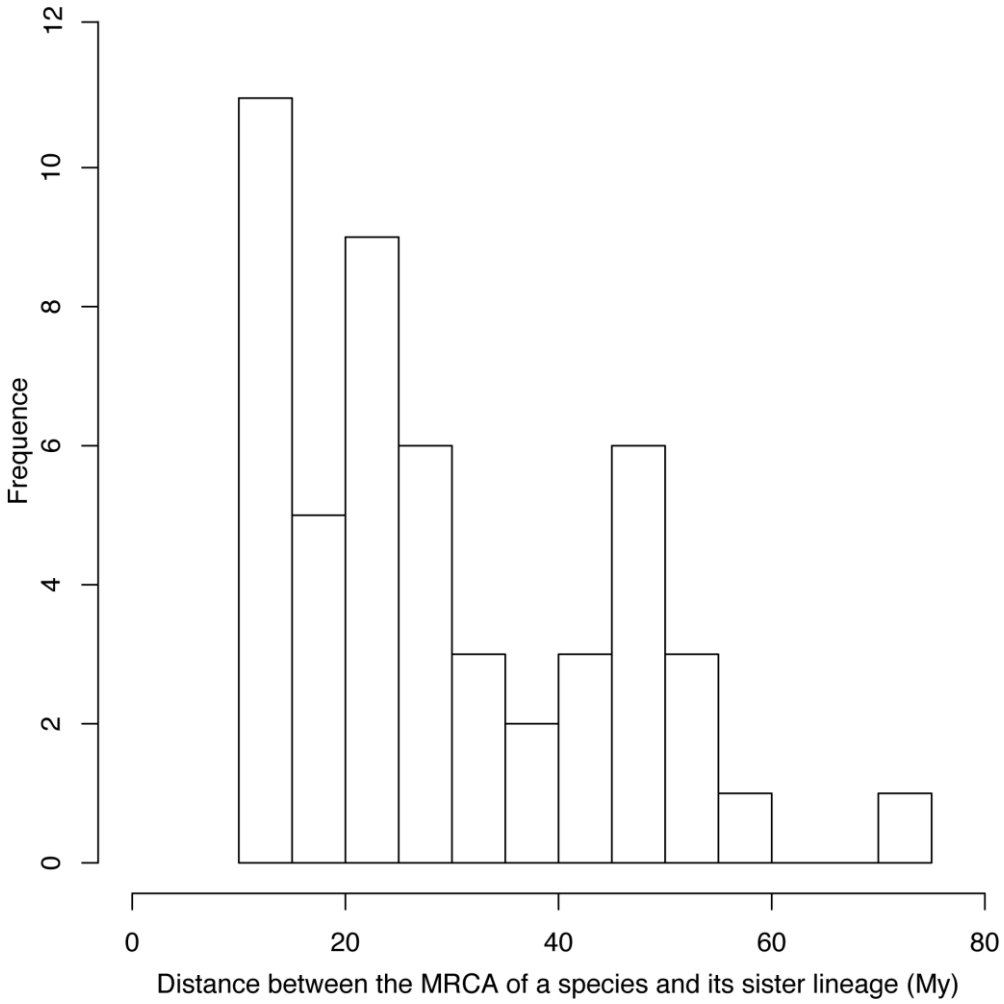


Figure 3. – Distribution in the timing of terminal-lineages diversification, as a function of the age of splitting with the sister lineage. Age is in Million years ago (Mya). Occurrences are counted according to five-million-years temporal classes (ranging from 0 to 75 Mya).

517 Supplementary Material 1 – Regions amplified for each sample (Genbank accessions to be

518 provided).

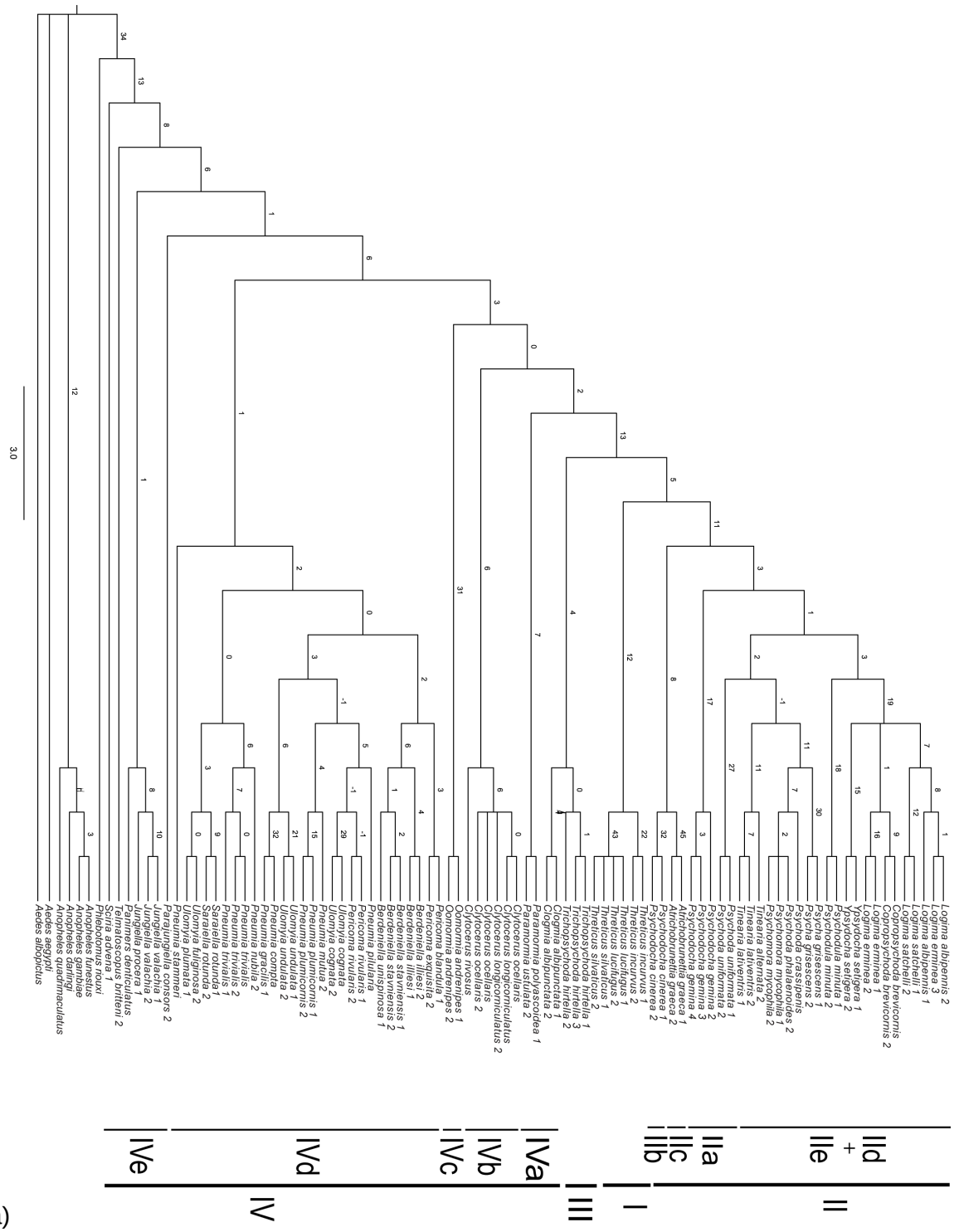
Sample	NADH-1	Cytochrome B
Berdeniella_illiesi1	-	yes
Berdeniella_illiesi2	-	yes
Berdeniella_stavniensis1	-	yes
Berdeniella_stavniensis2	-	yes
Berdeniella_unispinosa1	-	yes
Clogmia_albipunctata1	-	yes
Clogmia_albipunctata2	-	yes
Clytocerus_longicorniculatus	yes	yes
Clytocerus_longicorniculatus2	-	yes
Clytocerus_ocellaris	yes	yes
Clytocerus_ocellaris2	yes	yes
Clytocerus_rivosus	-	yes
Copropsychoda_brevicornis	yes	yes
Copropsychoda_brevicornis_2	yes	-
Jungiella_valachia1	yes	yes
Jungiella_valachia2	yes	yes
Logima_albipennis_1	-	yes
Logima_albipennis_2	yes	yes
Logima_erminea_1	yes	-
Logima_erminea_2	yes	-
Logima_erminea_3	-	yes
Logima_satchelli_1	-	yes
Logima_satchelli_2	-	yes
Attrichobrunettia_graeca1	-	yes
Attrichobrunettia_graeca2	-	yes
Oomormia_andrenipes_1	yes	yes
Oomormia_andrenipes_2	yes	yes
Pericoma_blandula1	yes	yes
Parajungiella_consors2	yes	yes
Paramormia_polyascoidea1	yes	-
Paramormia_ustulata2	yes	yes
Panimerus_denticulatus	-	yes
Pericoma_exquisita2	-	yes
Pericoma_rivularis1	yes	yes
Pericoma_rivularis2	-	yes
Pneumia_compta	-	yes
Pneumia_gracilis1	yes	yes
Pneumia_mutua2	-	yes
Pneumia_nubila2	yes	-
Pneumia_pilularia	-	-
Pneumia_plumicornis1	yes	yes
Pneumia_plumicornis2	yes	yes
Pneumia_stammeri	yes	yes
Pneumia_trivialis3	yes	yes
Pneumia_trivialis2	yes	-
Pneumia_trivialis1	yes	yes

Sample	NADH-1	Cytochrome B
Psychodocha_cinerea_1	-	yes
Psychodocha_cinerea_2	yes	yes
Psychodula_minuta_1	-	yes
Psychoda_crassipenis	-	yes
Psychodocha_gemina_2	-	yes
Psychodocha_gemina_3	-	yes
Psychodocha_gemina_4	-	yes
Psycha_grisescens_1	-	yes
Psycha_grisescens_2	yes	yes
Psychodula_minuta_2	yes	yes
Psychoda_phalaenoides_2	-	yes
Jungiella_procera1	-	yes
Psychoda_uniformata_1	yes	yes
Psychoda_uniformata_2	yes	yes
Psychomora_mycophila_1	yes	-
Psychomora_ mycophila_2	yes	yes
Saraiella_rotunda1	-	yes
Saraiella_rotunda2	yes	yes
Sciria_advena1	yes	yes
Telmatoscopus_britteni2	-	yes
Threticus_incurvus1	yes	yes
Threticus_incurvus2	-	yes
Threticus_lucifugus1	yes	yes
Threticus_lucifugus2	yes	yes
Threticus_silvaticus1	yes	yes
Threticus_silvaticus2	yes	yes
Tinearia_alternata_2	yes	yes
Tinearia_lativentris1	yes	-
Tinearia_lativentris2	yes	yes
Trichopsychoda_hirtella2	yes	yes
Trichopsychoda_hirtella3	yes	yes
Ulomyia_cognata	-	yes
Ulomyia_cognata2	-	yes
Ulomyia_fuliginosa_2	-	yes
Ulomyia_plumata1	-	yes
Ulomyia_undulata1	-	yes
Ulomyia_undulata2	yes	yes
Ypsydocha_setigera_1	yes	yes
Ypsydocha_setigera_2	yes	yes
Aedes albopictus	yes	yes
Anopheles funestus	yes	yes
Aedes aegypti	yes	yes
Anopheles darlingi	yes	yes
Anopheles quadrimaculatus	yes	yes
Anopheles gambiae	yes	yes
Phlebotomus riouxi	-	yes

519 Supplementary Material 2 – Maximum Parsimony topology with Decay Indexes (a) and

520 Maximum Likelihood phylogeny with bootstrap values (b).

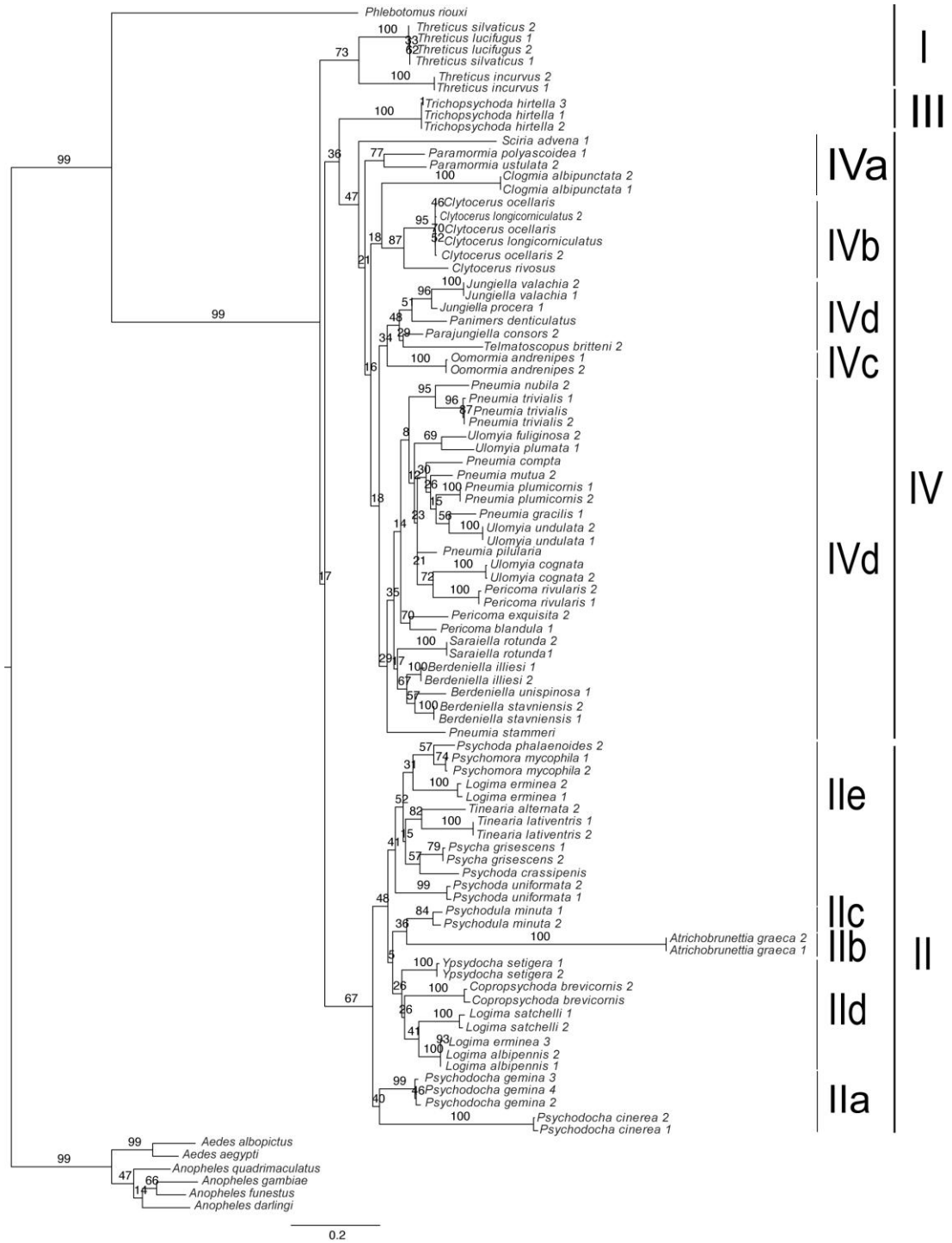
521



522 a)

523

b)


Phlebotomus riouxi
Threticus silvaticus 2
Threticus lucifugus 1
Threticus lucifugus 2
Threticus silvaticus 1
Threticus incurvus 2
Threticus incurvus 1
Trichopsychoda hirtella 3
Trichopsychoda hirtella 1
Trichopsychoda hirtella 2
Sciria advena 1
Paramormia polyascoidea 1
Paramormia ustulata 2
Clogmia albipunctata 2
Clogmia albipunctata 1
Clytocerus ocellaris
Clytocerus longicorniculatus 2
Clytocerus ocellaris
Clytocerus longicorniculatus
Clytocerus ocellaris 2
Clytocerus rivosus
Jungiella valachia 2
Jungiella valachia 1
Jungiella procera 1
Panimers denticulatus
Parajungiella consors 2
Telmatoscopus britteni 2
Oomormia andrenipes 1
Oomormia andrenipes 2
Pneumia nubila 2
Pneumia trivialis 1
Pneumia trivialis
Pneumia trivialis 2
Ulomyia fuliginosa 2
Ulomyia plumata 1
Pneumia compta
Pneumia mutua 2
Pneumia plumicornis 1
Pneumia plumicornis 2
Pneumia gracilis 1
Ulomyia undulata 2
Ulomyia undulata 1
Pneumia pilularia
Ulomyia cognata
Ulomyia cognata 2
Pericoma rivularis 2
Pericoma rivularis 1
Pericoma exquisita 2
Pericoma blandula 1
Saraiella rotunda 2
Saraiella rotunda1
Berdeniella illiesi 1
Berdeniella illiesi 2
Berdeniella unispinosa 1
Berdeniella stavniensis 2
Berdeniella stavniensis 1
Pneumia stammeri
Psychoda phalaenoides 2
Psychomora mycophila 1
Psychomora mycophila 2
Logima erminea 2
Logima erminea 1
Tinearia alternata 2
Tinearia lativentris 1
Tinearia lativentris 2
Psycha grisescens 1
Psycha grisescens 2
Psychoda crassipenis
Psychoda uniformata 2
Psychoda uniformata 1
Psychodula minuta 1
Psychodula minuta 2
Atrichobrunettia graeca 2
Atrichobrunettia graeca 1
Ypsydocha setigera 1
Ypsydocha setigera 2
Copropsychoda brevicornis 2
Copropsychoda brevicornis
Logima satchelli 1
Logima satchelli 2
Logima erminea 3
Logima albipennis 2
Logima albipennis 1
Psychodocha gemina 3
Psychodocha gemina 4
Psychodocha gemina 2
Psychodocha cinerea 2
Psychodocha cinerea 1
Aedes albopictus
Aedes aegypti
Anopheles quadrimaculatus
Anopheles gambiae
Anopheles funestus
Anopheles darlingi
0.2
I
III
IVa
IVb
IVd
IVc
IV
IVd
IIe
IIc
IIb
II
IId
IIa

Chapter three

Variation in the proportion of flower visitors of Arum maculatum along its distributional range in relation with community-based climatic niche analyses

Anahí Espíndola, Loïc Pellissier
and Nadir Alvarez

In press OIKOS

Variation in the proportion of flower visitors of *Arum maculatum* along its distributional range in relation with community-based climatic niche analyses

Anahí Espíndola, Loïc Pellissier and Nadir Alvarez

A. Espíndola (maria.espindola@unine.ch), Laboratory of Evolutionary Entomology, Inst. of Biology, Univ. of Neuchâtel, Emile-Argand 11, CH-2000 Neuchâtel, Switzerland. – L. Pellissier and N. Alvarez, Dept of Ecology and Evolution, Univ. of Lausanne, Biophore Building, CH-1015 Lausanne, Switzerland.

Because species–specific interactions between plants and insects require considerable physiological adaptations to establish and be maintained through time and space, highly specialized interactions are rare in nature. Consequently, even if some one-to-one interactions might appear locally specialized, additional partners may be involved at a wider scale.

Here, we investigate the geographical constancy in the specificity level of the specialized lure-and-trap pollination antagonism involving the widespread European *Arum maculatum* and its associated Psychodid pollinators. Until now, studies concurred in demonstrating that one single insect species, *Psychoda phalaenoides*, efficiently cross-pollinated plants; researches were, however, performed locally in western Europe. In this study we characterize for the first time the flower visitors' composition at the scale of the distribution range of *A. maculatum* by intensively collecting plants and insects throughout the European continent. We further correlate local climatic characteristics with the community composition of visiting arthropods.

Our results show that flowers are generally visited by *P. phalaenoides* females, but not over the whole distribution range of the plant. In some regions this fly species is less frequent or even absent and another species, *Psycha grisescens*, becomes the prevailing visitor. This variability is geographically structured and can be explained by climatic factors: the proportion of *P. grisescens* increases with higher annual precipitations and lower precipitations in the warmest trimester, two characteristics typical of the Mediterranean zone. Climate thus seems driving the specificity of this interaction, by potentially affecting the phenology of one or both interacting species, or even of volatile and heat production in the plant. This result therefore challenges the specificity of other presumably one-to-one interactions covering wide distribution ranges, and provides an example of the direct effect that the abiotic environment can have on the fate of plant–insect interactions.

In its early definition, the coevolutionary theory predicted that adaptive responses and speciation patterns in organisms were strongly influenced by one or a few interacting species (Ehrlich and Raven 1964). This has led to a simple view of coevolution, in which the evolutionary history of interacting organisms mostly occurs in a strict framework of coadaptation and cocladogenesis, as formerly proposed in the case of figs and fig-wasps (Ramirez 1970, Jousselin et al. 2008). Establishing highly specific relationships would thus represent an evolutionary stable strategy, since it would imply a rise in the efficiency of exploitation, and would assure the constancy through time of the resource(s) exploited (Bronstein 2009). However, such cases of species-specific interactions between plants and insects are more the exception than the rule (Hoeksema and Bruna 2000) and the view of strict coevolution evolved towards a more complex and diffuse coevolutionary concept taking place at the community level (Stanton 2003). Reasons for the scarcity of one-to-one interactions include the fact that specific biological associations require the development of particular ecological and physiological features in at least one of the interacting lineages (Brodie and Ridenhour 2002), which might entail substantial energy investments. Moreover, specialization could also cause a loss of efficiency in the exploitation of other non-specific associated species (Thompson 1994), which potentially makes these interactions sensitive to the loss of their specific partners. Thompson's theory on the geographic mosaic of coevolution (Thompson 2005) also predicts specialized interactions to be scarce in communities characterized by varying or disturbed environments: because biological interactions are expected to change along the distributional gradient, partner variations and multiple associations (McKinney 1997, Débarre and Gandon 2010) could be favoured and specialization might be then evolutionarily non-adaptive. Even if Thompson's view is sound, investigating simultaneously environmental variations and interactions is, in practice, a complex challenge. As a consequence, the environment-interaction relationship has been

only analyzed in a few multi-partners case-studies (reviewed by Laine 2009). In contrast, specific interactions provide simple and adequate natural laboratories for this type of studies, because of the restricted number of interacting species.

The genus *Arum* (Araceae) is characterized by the display of specialized antagonistic pollination interactions with arthropods, which have been documented for all but one species (Gibernau et al. 2004, Diaz and Kite 2006, Lobin et al. 2007): insects visit the plant but no service or resource is provided in exchange for pollination. One of the most studied pollination systems in the genus is that of lords-and-ladies *Arum maculatum*, a widely distributed European species (Boyce 2006, Espíndola et al. 2010). Its monoecious and protogynous inflorescences are formed by a spadix surrounded by a spathe, a characteristic feature in Araceae (Mayo et al. 1997). The lower part of the spathe is narrowed and encloses the fertile flowers, forming the floral chamber. During anthesis the chamber opening is covered by modified male flowers, the staminodes (Boyce 1993), which thus close the only aperture. In combination to these morphological features, plants of this species present a complex combination of heating cycles and odour production, which help lure small coprophillous Psychodid flies. Duped flies are attracted to the smelling flower and, after landing on the oily spathes, slip into the floral chamber. Because of the staminodes, they cannot fly away, being thus trapped for several hours. While trying to flee they move in the chamber and will – if they had been recently trapped by another *Arum* – passively pollinate the receptive female flowers. After the female phase, pollen release takes place and the trapped hairy flies are covered with it. Once pollen is liberated, staminodes wither, walls lose their oily nature and the captive insects are set free (Lack and Diaz 1991).

Among the *Arum* species known to show lure-and-trap pollination, *A. maculatum* has called the attention of ecologists for decades notably because it appears to display a highly specific pollination biology (Beck 1983, Lack and Diaz 1991, Diaz and Kite 2002). Whereas other *Arum* species are pollinated by a guild of visiting insects (Gibernau et al. 2004), investigations concurred in demonstrating that *A. maculatum* relied only on the fly species *Psychoda phalaenoides* (Diptera: Psychodidae) (Diaz and Kite 2002). Even if researchers have also identified other species (mainly *Smittia pratorum* (Diptera: Chironomidae) and *Psycha grisescens* (Diptera: Psychodidae)) trapped in floral chambers, these were present in very small proportions compared to *P. phalaenoides*, and because they were carrying significantly less pollen grains than the former it was assumed that their occurrence in the flower was accidental (Diaz and Kite 2002). Since at the several locations studied in western Europe the species composition of visiting insects was similar, and because in all sites *P. phalaenoides* was by far the most frequent visitor, it became tacitly accepted that the interaction between the plant and *P. phalaenoides* was species-specific (Gibernau et al. 2004) and constant all over the *A. maculatum* distribution range. However, the plant is distributed across a much wider area than that studied since it covers a large part of the occidental Palearctic from Spain to Denmark and from Ireland to Bulgaria (Boyce 2006) and thus potentially spans a wide diversity of climatic conditions and environments. The theory of the geographic mosaic of coevolution would therefore predict that specialization should not be maintained throughout the plant's distribution range, and that the identity of pollinators might change in space, along environmental conditions. In order to test such predictions, we aim in the present work at: 1) evaluating the visitors composition of *A. maculatum* inflorescences with an unprecedented coverage of the whole plant's distribution range; 2) analyzing to what extent the putative spatial variation in visitors identities is correlated with environmental variations associated with climatic factors.

Material and methods

Sampling and insect identification

Fifty-six populations of *Arum maculatum* were sampled across the whole plant's distribution range (Supplementary material Appendix 1) during springs 2006, 2007 and 2008. To avoid biases related either to regional differences in the sampling effort or to local inter-annual variation, populations were sampled when possible based on a 200km grid (not shown), and all main European regions (i.e. central and northern Europe, Balkans, Italian Peninsula) were visited all years. Plant populations were sampled during blooming period and depending on populations sizes, up to ten floral chambers per population were randomly selected and collected (Supplementary material Appendix 1). Floral chambers and their visitors were independently kept in 70% ethanol. Insects were identified when possible to species with the help of entomological keys (Chinery 1976, Withers 1988, Ježek 1990, Oosterbroek 2006). Arthropods were counted, and proportions were calculated per population, per floral chamber and for the totality of the samples.

Spatial and environmental structure in the composition of visitors

Spatial variation of the floral chamber fauna was analyzed using ArcMap 9.3 (ESRI, Redlands CA, USA) by plotting and examining the compositions of all sampled visitors. To test for climatic factors influencing the composition of visitors found in the floral chambers, we performed a canonical correspondence analysis (CCA) (ter Braak 1986) using the library ade4 (Thioulouse et al. 1997) on R 2.10.0 (R Development Core Team 2010). The CCA is designed to extract synthetic environmental gradients from ecological data sets, by describing the differential habitat preferences (i.e. from a climatic point of view in our study) of taxa via an ordination diagram. Climatic values were extracted for each sampled point from seven WORLDCLIM layers at a resolution of 2.5 arc-minutes (Hijmans et al. 2005). Layers were chosen based on their ability to describe basic general characteristics of the environment (mean annual and seasonal temperatures and precipitations). Finally, the CCA was calculated on visitor abundances at different taxonomic levels (classes, orders, families, different species of Psychodids; Supplementary material Appendix 1).

Simultaneously, in order to identify the climatic factors explaining the variation in abundances of the two most

frequent and constant visiting species (i.e. *Psychoda phalaenoides* and *Psycha grisescens*), we calculated their proportion at each population, and adjusted a general linear model (GLM) with a binomial distribution considering the climatic variables previously extracted. Finally, the significance of the parameters estimated was tested by a Wald-z statistic, as implemented in the function glm in R 2.10.0.

Results

General composition of visitors

Two hundred and eighty-four floral chambers were collected from the 56 visited locations, containing a total of 11 036 arthropods. The mean number of arthropods trapped per population and per floral chamber were 187.05 (SE = 53.22) and 38.58 (SE = 8.03), respectively.

Absolute overall proportions were as follows: arthropods belonged to classes Arachnida (5.53%) and Insecta (94.47%). The totality of Arachnida was represented by the subclass Acarina. Insecta were composed of orders Diptera (92.68%), Collembola (1.21%), Hemiptera (0.38%) and Coleoptera (0.19%). The far most important proportion of Diptera was formed by the family Psychodidae (87.49%) in which main visitors were females of *Psychoda phalaenoides* (84.27%), followed by *Psycha grisescens* from both genders (2.06% females and 0.70% males); these two species together represented 87.03% of all arthropods trapped in the flower chambers. The remaining Psychodid species were very little represented and accounted for 0.46% of the total sampling.

Based on these results and in order to identify general patterns in the composition of visiting insects, sampled locations were classified according to the relative numbers of the two most abundant species *P. phalaenoides* and *P. grisescens*: locations comprising more *P. phalaenoides* than *P. grisescens* were defined as 'p' sites, while locations comprising more *P. grisescens* than *P. phalaenoides* were considered as 'g' sites (Supplementary material Appendix 2). As shown in Fig. 1, sites 'p' and 'g' not only differed in the relative abundance of the two most frequent species but also in the composition and overall abundance of visitors. In 'p' sites, the composition was strongly dominated by females of *P. phalaenoides*, with very small proportions of other groups, and a particularly high number of visitors (>70) in about half of them. In 'g' sites, the number of visitors was low (a maximum of 77), both males and females of *P. grisescens* were dominant, but other insect groups were also found in non-negligible proportions (Supplementary material Appendix 1).

Geographic variation in the composition of visitors

Figure 2 shows the variation in the composition of floral chamber visitors all over the plant distribution. *Psychoda phalaenoides* is almost found at every location (in blue on Fig. 2) but in varying proportions. Populations in the southern edge of the sampled area include a more diverse fauna, mainly represented by *P. grisescens* (red and pink portions in Fig. 2a). Moreover, when pies are resized proportionally to the total number of visitors (Fig. 2b), a pattern appears, with populations from the south being also less visited than those from the north, and this, irrespective of the year of collection (data not shown, but see Supplementary material Appendix 1).

Influence of climate on visitors' composition

The composition of the floral chambers' fauna seems to be mainly determined by the first CCA axis (64.95% variance explained), with *P. phalaenoides,* Acarina and *P. setigera* showing an opposite ordination to that of the other groups (Fig. 3a). It is worth noting that the correlation between the abundances of the two most commonly found species (i.e. *P. phalaenoides* and *P. grisescens*) is strongly negative. Regarding climatic variables (Fig. 3b), 'p'-type and 'g'-type populations are differentiated by temperatures, which are negatively correlated to the amount of precipitations: 'g'-type populations are associated with higher temperatures and lower rain rates than are 'p'-type.

The GLM shows that proportions of the two main visitors *P. phalaenoides* and *P. grisescens* are explained by annual precipitations ($p = 0.023$; Table 1) and precipitations during the warmer quarter ($p = 0.007$; Table 1). The proportion of *P. grisescens* increases with higher annual precipitations and lower precipitations during the warmer quarter. The opposite effect is observed for *P. phalaenoides*.

Discussion

A more complete insight on the composition of *A. maculatum* pollinators

Species-specific interactions are rare, because their persistence strongly relies on tight and durable coevolutionary forces. Such conditions are particularly difficult to maintain

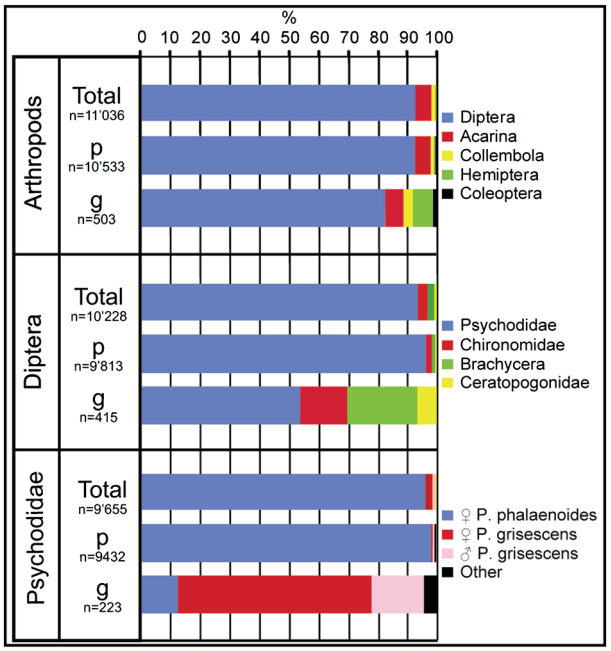


Figure 1. Proportions of Arthropods, Diptera and Psychodidae found in floral chambers. Total numbers according to the two pollinators' community categories are indicated (i.e. 'p': populations harbouring more *P. phalaenoides* than *P. grisescens*; 'g': populations harbouring more *P. grisescens* than *P. phalaenoides*; see text). Legend is shown at the right of the corresponding bars.

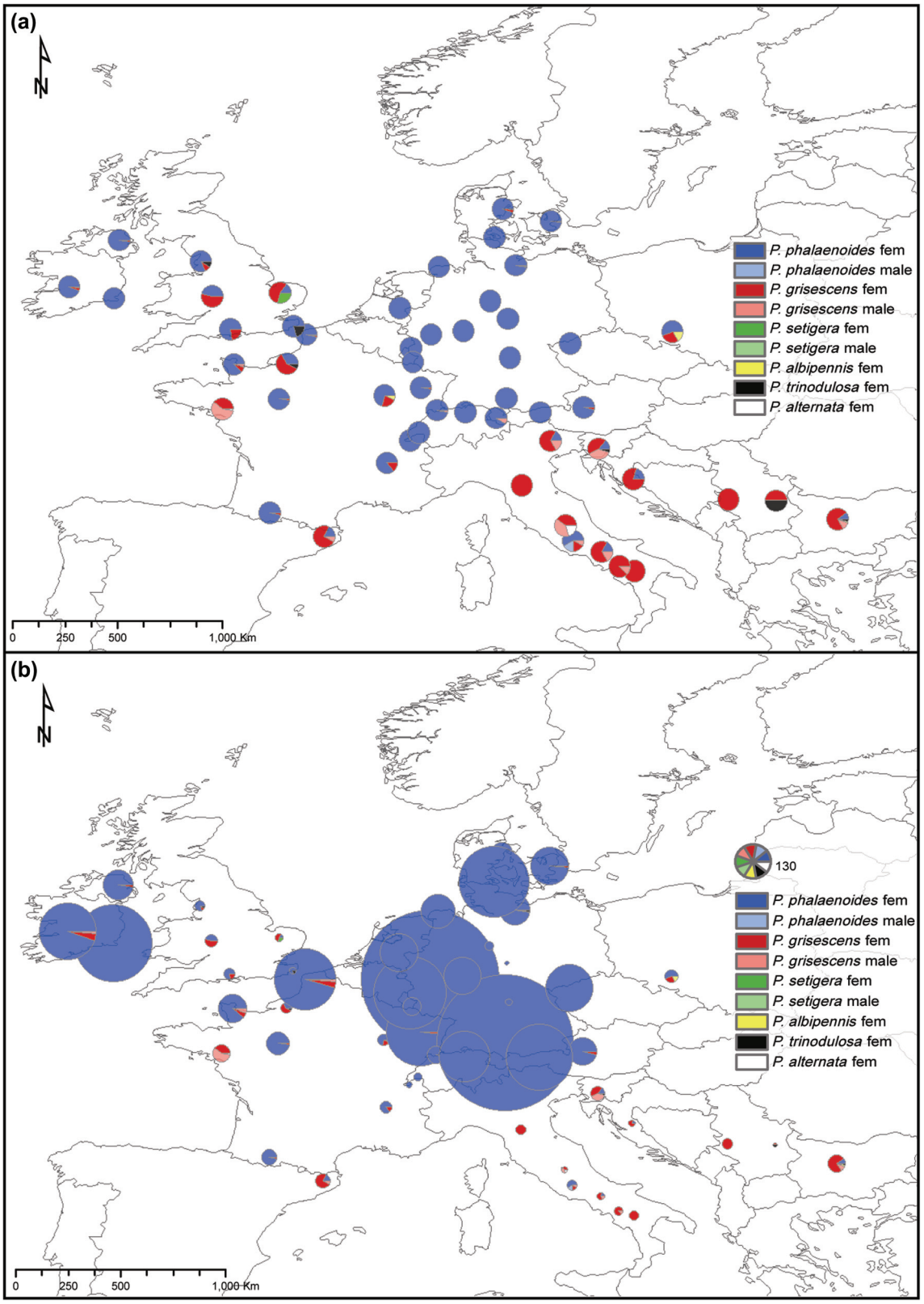


Figure 2. Distribution and composition of Psychodids found in floral chambers. Colours indicate species and sex. (a) proportions of Psychodids per population. (b) proportions resized in function of total number of trapped Psychodids (for a size reference, refer to map legend).

in species with wide distributions, in which environmental pressures are highly heterogeneous across their distribution ranges (Thompson 1994). Accordingly, our results indicate that the reported consistency in the strong and specific interaction between *A. maculatum* and its main pollinator is not preserved throughout the whole distribution range of the plant: both the total number of visitors and their composition are structured spatially (Fig. 2). Overall, there was

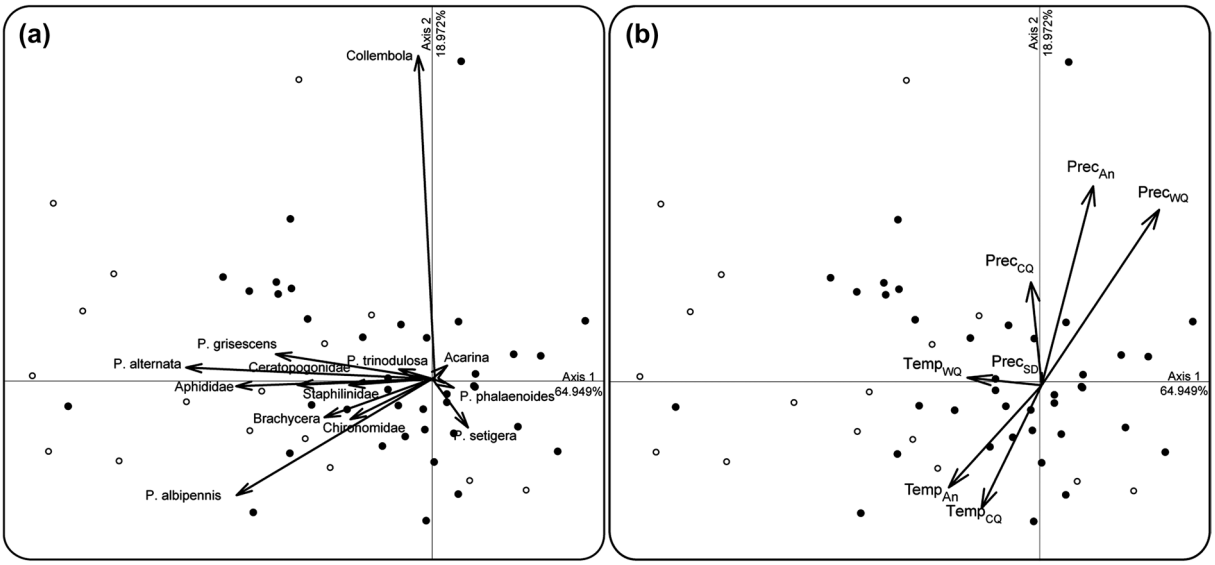


Figure 3. Canonical correspondence analysis for (a) the number and composition of arthropods found in floral chambers and (b) climatic variables. Variables are indicated on each arrow; direction of arrows indicates the sense of correlations. Filled circles represent type-'p' populations, while open circles refer to type-'g' populations. Percentage of variance explained is shown on axes. Refer to Table 1 for explanations on variables in (b).

a large variation in the abundance of *P. phalaenoides* specimens at each sampled site (ranging from zero to more than a thousand, Fig. 2b). Populations harbouring high numbers of *P. phalaenoides* females were mainly found in central and northern Europe, whereas those hosting a lower number of visitors – generally more diverse and mainly composed of *P. grisescens* males and females (82.96%) – were found in the Mediterranean area as well as in northern France (Fig. 2, Supplementary material Appendix 2). The similar behaviours, morphologies and ecologies shared by *P. phalaenoides* and *P. grisescens* probably attest to the ability of the latter in transporting *A. maculatum* pollen. However, the pollination effectiveness of *P. grisescens* should be investigated in these populations before unambiguously confirming its identity as a real pollinator. Because *A. maculatum* can also reproduce by clonal means (Boyce 1993), further population genetics analyses in the plant all over its southern range of distribution could provide information on its level of cross-pollination and consequently, on the ability of *P. grisescens* to efficiently transfer pollen from one plant to another. Finally, other arthropods captured by *A. maculatum* were less abundant and were not structured geographically (Fig. 2). They appear to be more casual and accidental visitors than reliable pollen vectors, but this does not exclude their potential importance in cross-pollination when main visitors lack.

Table 1. Estimates, z-values and associated p-values of the GLM adjusted to the proportion of *P. grisescens* and *P. phalaenoides*. *: $p < 0.05$.

Climatic variable	Estimate	z-value	p-value
Intercept	–4.793	–0.74	0.460
Mean annual temperature (Temp$_{Ann}$)	0.131	1.026	0.305
Annual precipitations (Prec$_{Ann}$)	0.024	2.27	0.023*
Temp. warmest quarter (Temp$_{WQ}$)	–0.040	–0.514	0.607
Temp. coldest quarter (Temp$_{CQ}$)	–0.084	–1.487	0.137
Precip. SD (Prec$_{SD}$)	0.062	1.155	0.248
Precip. warmest quarter (Prec$_{WQ}$)	–0.052	–2.79	0.005*
Precip. coldest quarter (Prec$_{CQ}$)	–0.048	–1.801	0.072

Variation in the attracted insect genders

Our study points out that *A. maculatum* flowers capture mainly females of *P. phalaenoides*, but both sexes of *P. grisescens* (Fig. 1). This result is in agreement with the findings of Diaz and Kite (2002), who described a similar trend in southern England. It is therefore likely that the odour cues released by the plant are different or are interpreted differently by the two species. Whereas *P. phalaenoides* females are attracted by *A. maculatum* odours that mimic their laying substrate (i.e. cow-dung; Kite 1995), one could imagine that this same odour also represents a mating site for *P. grisescens* (at least in males). This would be plausible since cases in which insects mate at their laying sites are not rare (Crowson 1981, Petit et al. 2007).

The effect of the environment on plant–insect interactions

Our results show that the composition of arthropods found in floral chambers of *A. maculatum* is correlated with climatic variables (Fig. 3, Table 1). The two most abundant visitors (i.e. *P. phalaenoides* and *P. grisescens*) were negatively correlated in the CCA, indicating that there is a strong differentiation of populations visited by one or another species in response to environmental conditions (Fig. 3a): populations dominated by *P. grisescens* (type 'g', open circles in Fig. 3) were mainly clustered in regions displaying higher summer temperatures combined with lower rain rates. This pattern is also confirmed by the GLM results, in which annual precipitations and precipitations during the warmest quarter were significantly correlated (positively and negatively, respectively) with the relative proportions of *P. grisescens* and *P. phalaenoides*.

It appears that precipitation-related factors modify the visitors' composition and do therefore affect the dynamics of the interaction between *A. maculatum* and its pollinators. These results tend to show that predictions from the theory of the geographic mosaic of coevolution (i.e. variation of the interaction in a variable environment) apply to the case of *A. maculatum*. This is, to our knowledge, one of the few studies (see also Toju 2008) in which the expectations of Thompson's theory match the observed pattern in the context of interacting species, and the first time it is demonstrated in plant–insect antagonistic interactions beneficial to the plant.

The proximal factors underlying this effect might be explained by three non-mutually exclusive hypotheses. First, because most 'g' sites were found in the Mediterranean region – which is characterized by rain regimes different from those of temperate Europe, with at least one marked dry season (Quézel 1985) – water stress might influence the plant phenology, leading it to flower at different time periods, uncoupled with the life cycle of *P. phalaeonoides*. As a consequence, it is possible that when Mediterranean *A. maculatum* blossom, their main pollinator *P. phalaenoides* is not present yet/anymore. Second, differences in precipitations could have an impact on the physiology of volatile and heat production, which are known to be both energetically costly (Minorsky 2003) and a key-factor for the attraction of pollinators in this system. Since changes in rain regimes and rates can put plants under water stress and modify their metabolism (Gouinguene and Turlings 2002, Baldwin et al. 2006), one could imagine that *A. maculatum* plants from sites visited mainly by *P. grisescens* present different odour/heat patterns than those visited mainly by *P. phalaenoides*, the latter relying on a much more constant water supply. These differences could make 'g'-type populations less successful than 'p'-type populations in terms of attraction of the most efficient pollinator (i.e. *P. phalaenoides*), this would also be in agreement with both their lower number of visitors and higher arthropod species diversity. Finally, it is also possible that changes in water resources modify the insects' life cycles, limiting their phenological overlap with plants blossom. This seems however unlikely, since faunistic records indicate that both fly species are always present long before and after the flowering period, and over a larger distribution range than that of the plant (Ježek and Barták 2000, Fauna Europaea 2004). An alternative would be that even if *P. phalaenoides* is known to occur in the Mediterranean region, this area constitutes a marginal habitat of lower suitability, in which the species is rather rare compared to more northern habitats. In order to test this hypothesis, further intensive entomological collections should be performed all over the insect's distribution range.

Conclusion

In this study we show that arthropod visitors of *A. maculatum* are mainly restricted to the two Psychodids *P. phalaenoides* and *P. grisescens*. Overall, as previously observed at a more limited geographical scale (Beck 1983, Diaz and Kite 2002), *P. phalaenoides* females are the most abundant visitors of *A. maculatum* flowers. However, the abundance and composition of the fauna captured in floral chambers strongly varies across the distribution range of the plant, with *P. grisescens* becoming notably more abundant in the Mediterranean region. As a consequence, diversity and abundance of arthropods are geographically structured (without any notable inter-annual effect), indicating that all over its distribution range the plant might exploit at least two species and not a single one, as previously assumed. Our results show that variation in environmental conditions, particularly in rain regimes, is correlated with changes in the visitors' composition. This climatic effect could be explained by the role played by water stress in modifying the phenology of either the plant or the insect, or the physiology of volatile and heat production in the plant. Precipitation-related factors therefore affect not only the visitors' composition of the plant, but also the specificity of the pollination interaction.

This study provides evidences confirming predictions of the theory of the geographic mosaic of coevolution, directly demonstrating that environmental variation does have an impact on antagonistic interactions. It finally points out the importance of applying large-scale surveys when challenging the existence of putative one-to-one interactions.

Acknowledgements – The authors would like to thank Dr. Z. Kącki and Dr. Z. Dajdok (Univ. of Wrocław, Poland), Dr. M. Gibernau (UMR CNRS, Univ. Paul Sabatier; France), Dr. J. Ježek (Natl Mus. of Prague; Czech Republic), Dr. F. Conti (Centro Ricerche Floristiche dell'Appennino, Italy), Dr. A. Diaz (Bournemouth Univ., England), N. Revel, Dr. Y. Triponez and P. Zryd (Univ. of Neuchâtel, Switzerland) for their precious assistance during sampling and data analysis. We are grateful to the journal editor A. Traveset, M. Méndez and A. Diaz for constructive comments on a previous version of this manuscript. This study was funded by the Swiss National Science Foundation (project no. 3100A0-116778).

References

Baldwin, I. T. et al. 2006. Volatile signaling in plant–plant interactions: "Talking trees" in the genomics era. – Science 311: 812–815.

Beck, I. 1983. Blütenökologie Untersuchungen am Aronstab (*Arum maculatum* L.). Spezielle Botanik. – Univ. of Tübingen, p. 123.

Boyce, P. C. 1993. The genus *Arum*. – R. Bot. Gard. Kew.

Boyce, P. C. 2006. *Arum* – a decade of change. – Aroideana 29: 132–137.

Brodie, E. D. and Ridenhour, B. J. 2002. Reciprocal selection at the phenotypic interface of coevolution. – Integr. Compar. Biol. 42: 1202–1202.

Bronstein, J. L. 2009. The evolution of facilitation and mutualism. – J. Ecol. 97: 1160–1170.

Chinery, M. 1976. Les insectes d'Europe. – Elsevier Séquoia.

Crowson, R. A. 1981. The biology of the Coleoptera. – Academic Press.

Débarre, F. and Gandon, S. 2010. Evolution of specialization in a spatially continuous environment. – J. Evol. Biol. 23: 1090–1099.

Diaz, A. and Kite, G. C. 2002. A comparison of the pollination ecology of *Arum maculatum* and *A. italicum* in England. – Watsonia 24: 171–181.

Diaz, A. and Kite, G. C. 2006. Why be a rewarding trap? The evolution of floral rewards in *Arum* (Araceae), a genus characterized by saprophilous pollination systems. – Biol. J. Linn. Soc. 88: 257–268.

Ehrlich, P. R. and Raven, P. H. 1964. Butterflies and plants: a study in coevolution. – Evolution 18: 586–607.

Espíndola, A. et al. 2010. New insights into the phylogeny and biogeography of *Arum* L. (Araceae): unravelling its evolutionary history. – Bot. J. Linn. Soc. 163: 14–32.
Fauna Europaea. 2004. Fauna Europaea ver. 1.1. <www.faunaeur.org>. Accessed 10 January 2010.
Gibernau, M. et al. 2004. Pollination in the genus *Arum* – a review. – Aroideana 27: 148–166.
Gouinguene, S. P. and Turlings, T. C. J. 2002. The effects of abiotic factors on induced volatile emissions in corn plants. – Plant Physiol. 129: 1296–1307.
Hijmans, R. J. et al. 2005. Very high resolution interpolated climate surfaces for global land areas. – Int. J. Climatol. 25: 1965–1978.
Hoeksema, J. D. and Bruna, E. M. 2000. Pursuing the big questions about interspecific mutualism: a review of theoretical approaches. – Oecologia 125: 321–330.
Ježek, J. 1990. Redescriptions of nine common palaearctic and holarctic species of Psychodini End. (Diptera: Psychodidae). – Acta Entomol. Mus. Natl Pragae 43: 33–83.
Ježek, J. and Barták, M. 2000. Psychodidae. – Folia Fac. Sci. Nat. Univ. Masaryk. Brun. 104: 93–100.
Jousselin, E. et al. 2008. One fig to bind them all: host conservatism in a fig wasp community unraveled by cospeciation analyses among pollinating and nonpollinating fig wasps. – Evolution 62: 1777–1797.
Kite, G. C. 1995. The floral odour of *Arum maculatum*. – Biochem. Syst. Ecol. 23: 343–354.
Lack, A. J. and Diaz, A. 1991. The pollination of *Arum maculatum* L. – a historical review and new observations. – Watsonia 18: 333–342.
Laine, A. L. 2009. Role of coevolution in generating biological diversity: spatially divergent selection trajectories. – J. Exp. Bot. 60: 2957–2970.

Supplementary material (available online as Appendix O18937 at <www.oikos.ekol.lu.se/appendix>). Appendix 1–2.

Lobin, W. et al. 2007. A new *Arum* species *(*Areae, Araceae*)* from NE Turkey and Georgia. – Willdenowia 37: 445–449.
Mayo, S. J. et al. 1997. The genera of Araceae. – R. Bot. Gard. Kew.
McKinney, M. L. 1997. Extinction vulnerability and selectivity: combining ecological and paleontological views. – Annu. Rev. Ecol. Syst. 28: 22.
Minorsky, P. V. 2003. The hot and the classic. – Plant Physiol. 132: 25–26.
Oosterbroek, P. 2006. The European families of the Diptera. Identification, diagnosis, biology. – KNNV Publishing.
Petit, D. et al. 2007. On the mating and laying sites of *Uromenus brevicollis* ssp. *insularis* in Corsica (Ensifera, Tettigoniidae). – J. Orthoptera Res. 16: 145–151.
Quézel, P. 1985. Definition of the Mediterranean region and the origin of its flora. – Trends Ecol. Evol. 23: 564–571.
Ramirez, W. B. 1970. Host specificity of fig wasps (Agaonidae). – Evolution 24: 680–690.
Stanton, M. L. 2003. Interacting guilds: moving beyond the pairwise perspective on mutualisms. – Am. Nat. 162: S10–S23.
ter Braak, C. J. F. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. – Ecology 67: 1167–1179.
Thioulouse, J. et al. 1997. ADE-4: a multivariate analysis and graphical display software. – Stat. Comput. 7: 75–83.
Thompson, J. N. 1994. The coevolutionary process. – Univ. of Chicago Press.
Thompson, J. N. 2005. Coevolution: the geographic mosaic of coevolutionary arms races. – Curr. Biol. 15: 992–994.
Withers, P. 1988. Moth flies. Diptera: Psychodidae. – Dipterist Digest 4: 83.

Oikos

o18937
Espíndola, A., Pellissier, L. and Alvarez, N. 2010. Variation in the proportion of flower visitors of *Arum maculatum* along its distributional range in relation with community-based climatic niche analyses. – Oikos 000: 000–000.

Appendix 1

List of populations and visitors' composition.

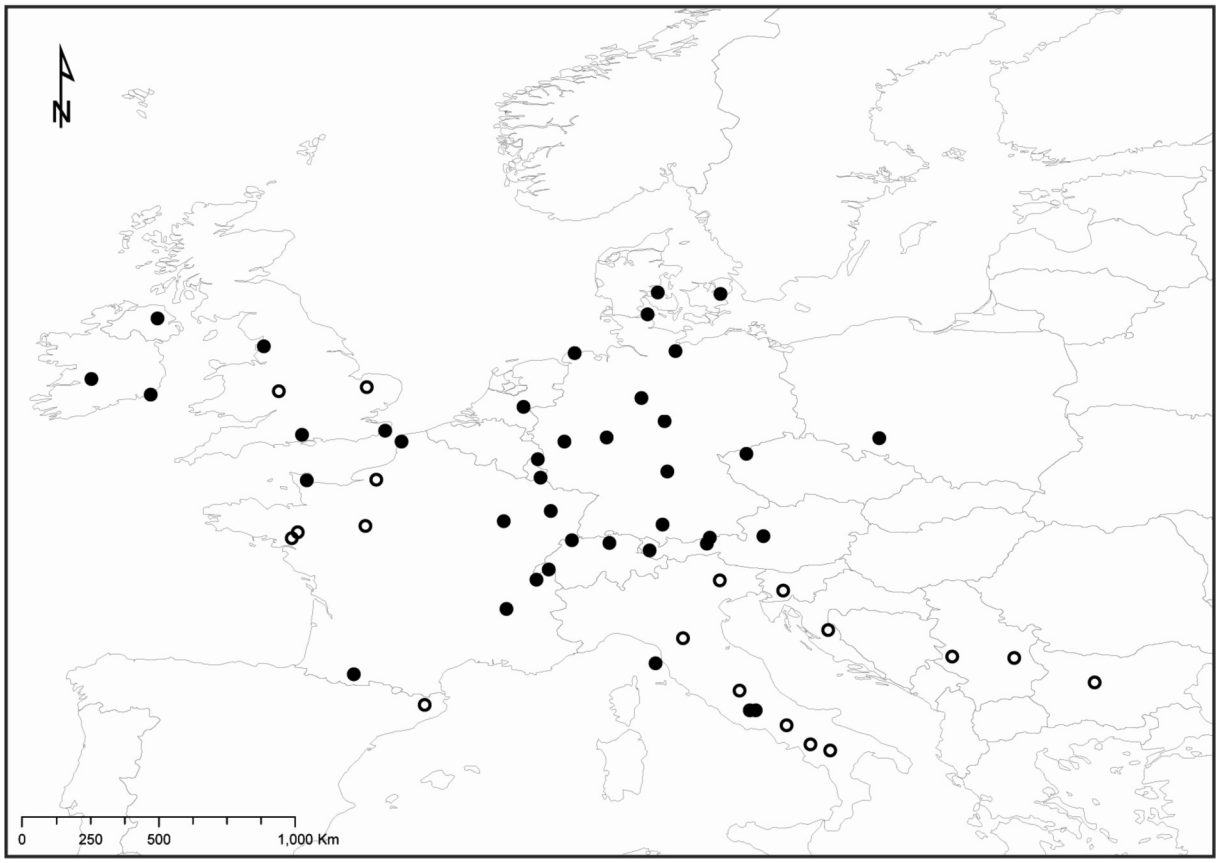
Location	Longitude	Latitude	Country	Altitude (m)	Year collection	Population type	Number of collected inflorescences	Diptera															Aphidae	Staphillinidae	Collembola	Acarina	Non-Diptera total	Total
								Psychodidae																				
								phalaenoides female	phalaenoides male	grisescens female	grisescens male	albipennis female	setigera female	setigera male	trinodulosa female	alternata female	Psychodidae total	Brachycera	Ceratopogonidae	Chironomidae	Non-Psychodidae total	Diptera total						
Aby	10° 10' 4.84"	56° 8' 35.81"	DK	0	2008	p	3	47	0	2	0	0	0	0	1	0	50	0	1	1	2	52	0	0	0	0	0	52
Acerno	15° 10' 6.74"	40° 46' 12.65"	IT	928	2007	g	4	0	0	7	1	0	0	0	0	0	8	0	0	4	4	12	0	1	0	0	1	13
Balzers	9° 53' 12.98"	47° 9' 3.99"	LICH	494	2006	p	4	146	0	1	14	0	0	0	0	0	161	1	2	0	3	164	0	0	0	20	20	184
Blankenheim	7° 5' 53.99"	50° 44' 2.00"	D	531	2006	p	5	2007	0	3	0	0	2	1	1	0	2014	2	3	2	7	2021	0	2	1	149	152	2173
Boc	13° 3' 56.05"	50° 20' 42.79"	POL	364	2008	p	5	236	0	0	2	0	0	0	0	0	238	1	1	0	2	240	0	1	1	11	13	253
Bouloire	0° 33' 40.86"	47° 58' 15.45"	F	115	2007	g	5	2	0	9	1	0	0	0	0	1	13	0	0	0	0	13	0	1	0	0	1	14
Bouloire	0° 33' 40.93"	47° 58' 15.53"	F	115	2008	p	5	52	0	1	0	0	0	0	0	0	53	0	0	3	3	56	1	0	3	0	4	60
Brzeg	17° 27' 3.24"	50° 51' 24.30"	POL	336	2008	p	5	10	0	5	0	3	0	0	0	0	18	6	17	5	28	46	0	0	0	0	0	46
Chaumont	5° 5' 41.10"	48° 6' 54.29"	F	296	2007	p	6	10	0	3	0	1	0	0	0	0	14	26	0	4	30	44	0	1	0	0	1	45
Chene-Bourg	6° 11' 7.19"	46° 11' 34.01"	F	0	2007	p	5	4	0	0	0	0	0	0	0	0	4	3	0	1	4	8	0	5	0	5	10	18
Chiflik	24° 31' 35.72"	42° 48' 26.53"	BUL	856	2008	g	7	3	0	23	3	1	0	0	1	0	31	1	0	0	1	32	0	0	2	1	3	35
Churchtown	-2° 47' 12.70"	53° 52' 52.36"	ENG	3	2007	p	7	9	0	1	0	0	0	0	1	0	11	6	0	8	14	25	0	1	0	1	2	27
Col de Roanza	12° 12' 35.60"	46° 10' 34.97"	IT	816	2008	g	4	1	0	4	1	0	0	0	0	0	6	2	6	0	8	14	0	0	1	0	1	15
Conteville	1° 44' 19.39"	50° 44' 14.32"	F	60	2007	p	8	379	2	13	5	0	0	0	4	0	403	28	0	53	81	484	0	0	3	53	56	540
Cranwich	0° 36' 29.45"	52° 31' 37.16"	ENG	24	2007	g	8	1	0	4	0	0	2	0	0	0	7	18	0	9	27	34	1	0	0	0	1	35
Dewley	-2° 17' 30.48"	52° 23' 11.47"	ENG	35	2007	g	8	8	0	9	0	0	0	0	0	0	17	7	0	19	26	43	0	1	1	2	4	47
Dornum	7° 25' 52.21"	53° 39' 3.42"	D	0	2008	p	5	120	0	0	0	0	0	0	0	0	120	1	0	0	1	121	0	0	1	1	2	123
Duclair	0° 55' 12.58"	49° 29' 6.61"	F	-1	2007	g	7	4	0	7	0	0	0	0	1	0	12	30	2	19	51	63	0	3	0	11	14	77
Ecausseville	-1° 22' 27.88"	49° 27' 43.88"	F	0	2008	p	6	75	0	6	4	0	0	0	0	0	85	0	2	0	2	87	1	0	0	0	1	88
Ecausseville1	-1° 22' 27.80"	49° 27' 43.85"	F	-26	2007	p	5	21	1	0	0	0	0	0	2	1	25	5	0	38	43	68	0	0	1	1	2	70

Location	Longitude	Latitude	Country	Altitude (m)	Year collection	Population type	Number of collected inflorescences	Diptera														Aphidae	Staphillinidae	Collembola	Acarina	Non-Diptera total	Total
								Psychodidae																			
								phalaenoides female	phalaenoides male	grisescens female	grisescens male	albipennis female	setigera female	setigera male	trinodulosa female	alternata female	Psychodidae total	Brachycera	Ceratopogonidae	Chironomidae	Non-Psychodidae total	Diptera total					
Ecublens	6° 33' 58.93"	46° 31' 35.58"	CH	0	2006	p	2	6	0	0	0	0	0	0	0	0	6	1	0	0	1	7	0	0	0	0	0	7
Einhaus	10° 44' 4.99"	53° 42' 48.78"	D	21	2008	p	4	94	0	1	0	0	0	0	2	0	97	0	0	0	0	97	1	0	0	0	1	98
Eldgasen	9° 37' 31.22"	52° 10' 31.19"	D	101	2008	p	5	8	0	0	0	0	0	0	0	0	8	1	0	0	1	9	0	0	0	0	0	9
Epinal	6° 39' 16.99"	48° 27' 1.01"	F	400	2006	p	5	474	0	9	4	0	0	0	0	0	487	0	0	0	0	487	0	0	0	12	12	499
Ewijk	5° 44' 54.82"	51° 52' 34.68"	NL	21	2008	p	1	151	0	0	1	0	0	0	0	0	152	0	0	0	0	152	0	0	0	0	0	152
Gjeddesdal	12° 13' 50.30"	55° 36' 27.18"	D	32	2008	p	5	149	0	2	1	0	0	0	0	0	152	0	1	0	1	153	0	0	0	0	0	153
Gostilje	19° 50' 7.76"	43° 39' 20.19"	SRB	785	2008	g	5	0	0	12	0	0	0	0	0	0	12	1	8	0	9	21	0	0	1	0	1	22
Hoenggerberg	8° 33' 58.93"	47° 24' 29.23"	CH	0	2006	p	6	273	0	0	0	0	0	0	0	0	273	0	0	0	0	273	0	0	0	0	0	273
Hautes pyrenees	0° 10' 43.25"	43° 4' 38.35"	F	580	2007	p	1	24	0	1	0	0	0	0	0	0	25	0	0	0	0	25	0	0	0	0	0	25
Hohenweiler	10° 18' 56.01"	48° 0' 24.98"	A	486	2006	p	5	1972	0	1	0	0	1	1	6	0	1981	2	0	3	5	1986	0	0	0	139	139	2125
Jenne	13° 10' 55.88"	41° 52' 53.11"	IT	824	2007	p	2	7	2	2	1	0	0	0	0	0	12	1	0	3	4	16	1	0	0	0	1	17
Johnstown Castle	-6° 30' 14.83"	52° 17' 32.82"	IRL	-18	2007	p	6	631	0	3	0	0	0	0	0	0	634	0	0	0	0	634	0	0	0	1	1	635
Killaloe	-8° 26' 39.77"	52° 47' 40.56"	IRL	44	2007	p	7	322	0	14	5	0	0	0	0	0	341	5	2	20	27	368	0	0	0	12	12	380
La Mignonnais	-1° 51' 52.78"	47° 33' 2.38$"	F	16	2007	g	5	0	0	13	19	0	0	0	0	1	33	0	2	3	5	38	0	0	1	0	1	39
Lago Matese	14° 24' 20.27"	41° 24' 23.08"	IT	1098	2008	g	7	1	0	4	1	0	0	0	0	0	6	11	1	0	12	18	0	0	2	0	2	20
Maiche	7° 20' 22.99"	47° 29' 5.99"	F	613	2006	p	4	32	0	0	1	0	0	0	0	0	33	0	1	0	1	34	0	0	0	0	0	34
Matese	14° 24' 20.19'	41° 24' 23.00"	IT	1098	2007	g	3	0	0	1	1	0	0	0	0	0	2	2	0	3	5	7	0	0	0	0	0	7
Montese	10° 59' 1.35"	44° 15' 18.83"	IT	707	2007	g	5	0	0	10	0	0	0	0	0	0	10	0	0	0	0	10	1	0	2	0	3	13
Morro Reatino	12° 51' 26.71"	42° 32' 9.06"	IT	759	2007	g	3	0	0	2	2	0	0	0	0	1	5	0	0	1	1	6	0	0	0	0	0	6
Randalstown	-6° 17' 0.56"	54° 47' 40.09"	IRL	58	2007	p	6	94	0	3	0	0	0	0	0	0	97	4	1	0	5	102	0	0	1	1	2	104
Rifreddo	15° 49'	40° 34'	IT	1172	2007	g	4	0	0	10	0	0	0	0	0	0	10	0	2	1	3	13	0	0	0	0	0	13

	29.03"	20.46"																									
Romsey	-1° 31' 29.93"	50° 57' 51.26"	ENG	52	2007	p	8	10	0	3	0	0	0	0	0	0	13	1	2	20	23	36	2	0	0	1	3	39
Sankt Margrethen	11° 46' 32.66"	47° 22' 57.00"	A	527	2008	p	5	469	0	0	0	0	0	0	0	0	469	9	4	34	47	516	0	0	0	15	15	531
Sitno	13° 37' 59.81'	47° 36' 48.56"	POL	372	2008	p	5	26	0	2	0	0	0	0	0	0	28	0	0	0	0	28	0	0	40	0	40	68
								Diptera																			
									Psychodidae																		
Location	Longitude	Latitude	Country	Altitude (m)	Year collection	Population type	Number of collected inflorescences	phalaenoides female	phalaenoides male	grisescens female	grisescens male	albipennis female	setigera female	setigera male	trinodulosa female	alternata female	Psychodidae total	Brachycera	Ceratopogonidae	Chironomidae	Non-Psychodidae total	Diptera total	Aphididae	Staphillinidae	Collembola	Acarina	Non-Diptera total	Total
Sokobanja Vrh	21° 53' 15.18"	43° 36' 13.43"	SRB	844	2008	g	5	0	0	1	0	0	0	0	1	0	2	14	0	0	14	16	31	0	1	10	42	58
Sonderborg	9° 49' 13.94"	54° 55' 4.51"	DK	7	2008	p	10	528	0	0	0	0	0	0	0	0	528	0	3	2	5	533	0	1	10	10	21	554
St. Fargeau	3° 5' 31.38"	47° 37' 20.39"	F	226	2007	p	6	0	0	0	0	0	0	0	0	0	0	4	0	11	15	15	0	0	0	7	7	22
Steeg	13° 37' 59.81"	47° 36' 48.56"	A	513	2008	p	1	53	0	1	0	0	0	0	0	0	54	0	0	0	0	54	0	0	0	6	6	60
Steeg2	13° 37' 59.91"	47° 36' 48.63"	A	513	2008	p	2	79	0	3	0	0	0	0	0	0	82	0	0	0	0	82	0	0	40	6	46	128
Steinach	10° 28' 4.01"	49° 44' 35.99"	D	363	2006	p	5	5	0	0	0	0	0	0	0	0	5	0	0	0	0	5	0	1	0	0	1	6
Tintesmuehle	6° 13' 26.00"	50° 9' 21.99"	LUX	305	2006	p	4	550	0	0	0	0	1	0	5	0	556	2	8	5	15	571	1	4	6	117	128	699
Udbina	15° 45' 40.82"	44° 31' 52.61"	HVR	810	2008	g	7	1	0	4	0	0	0	0	0	0	5	1	6	0	7	12	0	0	3	0	3	15
Unec	14° 17' 9.92"	45° 49' 42.42"	SLO	525	2008	g	5	3	0	9	8	0	0	0	1	0	21	12	0	7	19	40	1	0	1	8	10	50
Verne	5° 11' 25.80"	45° 13' 36.30"	F	502	2007	p	1	12	0	2	0	0	0	0	0	0	14	0	0	0	0	14	0	0	0	0	0	14
Weilburg	8° 28' 32.02"	50° 53' 0.99"	D	158	2006	p	3	146	0	0	0	1	0	0	0	0	147	0	0	0	0	147	0	0	0	10	10	157
Worbis	10° 22' 55.13"	51° 25' 15.17"	D	370	2008	p	4	2	0	0	0	0	0	0	0	0	2	0	3	0	3	5	1	0	0	0	1	6
Yutz	6° 18' 14.00"	49° 33' 5.00"	F	225	2006	p	4	35	0	0	0	0	0	0	0	0	35	0	0	0	0	35	0	0	0	0	0	35

Appendix 2

Sampling distribution of *A. maculatum* visitors. Open circles: type-'g' populations; full circles: type-'p' populations.



Chapter four

Size-advantage model and pollination strategies in plants

Natacha Revel, Marc Gibernau,
Nadir Alvarez and Anahí Espíndola

To be resubmitted to *Evolutionary Ecology*

Size-advantage model and pollination strategies in plants

Natacha Revel[1], Nadir Alvarez[2], Marc Gibernau[3], Anahí Espíndola[1 4]

[1] Laboratory of Evolutionary Entomology, Institute of Biology, University of Neuchâtel, Emile-Argand 11, 2009 Neuchâtel, Switzerland.

natacha.revel@gmail.com

[2] Department of Ecology and Evolution, University of Lausanne, Biophore Building, CH-1015 Lausanne, Switzerland. nadir.alvarez@unil.ch

[3] CNRS-UMR 8172 Ecologie des Forêts de Guyane, BP 709, 97387 Kourou, France. Marc.Gibernau@ecofog.gf

[4] Corresponding author. Laboratory of Evolutionary Entomology, Institute of Biology, University of Neuchâtel, Emile-Argand 11, 2009 Neuchâtel, Switzerland.

maria.espindola@unine.ch.

Running title: Size-advantage model and pollination strategies

Keywords: size-advantage model, pollination, plant-insect interactions, *Arum*

Total word count: 4744

Word count Abstract: 298

Word count Introduction: 1914

Word count Material and Methods: 727

Word count Results: 614

Word count Discussion: 1122

Word count Acknowledgements: 69

Number of references: 55

Number of Figures: 4

Number of Tables: 1

Online appendix: 1

Abstract

The size-advantage model (SAM) is aimed at explaining the temporal variation of energetic investment on reproductive structures in long-lived hermaphroditic organisms. The theory proposes that an increase in the resources available to an organism induces a higher relative investment on the energetically most costly sexual structures. The model has been demonstrated to apply to both animals and plants, and different factors (e.g., habitat quality, capability of dispersion) appear to shape the benefit curve of the SAM expression.

In plants, pollination interactions are known to play a role in the evolution of flower features. Because the SAM directly concerns flower characters, pollinators are expected to have an influence on the application of the model. This hypothesis has however never been tested.

In the present work, we investigate whether the identity and diversity of pollinators can be used as a proxy to predict the application of the SAM in exclusive zoophilous plants. We present a new approach to understand the dynamics of the model and test it on pollination systems involving widespread central-European *Arum* (Araceae) species (*A. cylindraceum*, *A. maculatum* and *A. italicum*).

By identifying the species composition, abundance and spatial variation of arthropods trapped in the inflorescences we show that *A. cylindraceum* displays a generalist reproductive strategy, relying on the exploitation of a low number of dipterans, similarly to what is found in *A. italicum*, but contrasting with the pattern depicted in *A. maculatum*. Based on the simple model presented here, the application of the SAM is predicted for the first two and rejected for the last species, predictions being further confirmed by allometric measures.

This is to our knowledge the first time that this theory is both proposed and empirically tested in zoophilous pollination systems. Its application on other non-*Arum* systems is discussed, indicating its general biological importance.

Introduction

Energetic income is limited in all organisms and natural selection optimizes resource allocation strategies between growth/survival and reproduction according to physiological and environmental (biotic and abiotic) factors (Guillon et al. 2006; Lloyd and Bawa 1984; Policansky 1982). As a consequence, it has been proposed that in hermaphroditic and long-lived organisms the reproductive energetic allocation varies with the size of the organism. In general, when available resources increase, the size of an organism tends to augment, which favours the relative increase of the most costly sexual structures. This model –namely the Size-Advantage Model (SAM)– was first proposed (Ghiselin 1969) and tested (Warner 1988) in animals, but was later applied to dioecious and hermaphroditic plants (Lloyd and Bawa 1984), *e.g.*, in species within genera *Acer* L. (Sapindaceae) (Primack and McCall 1986), *Arisaema* Mart. (Araceae) (Vitt et al. 2003) and *Arum* L. (Araceae) (Méndez 2001).

According to the model, floral resource allocation, *i.e.*, the quantity of energy invested by a plant in the production of female *vs.* male sexual structures (Lloyd and Bawa 1984), depends on the energy available at a certain time. When only the amount of available resources varies, larger inflorescences -found in larger plants- are predicted to harbour a larger proportion of female flowers (i.e., the most energetically costly sexual structure) than smaller inflorescences. Since its first application, the comprehension of the SAM has been progressively enhanced by the inclusion of additional factors such as local sexual competition and differential gamete dispersion (Lloyd and Bawa 1984), wind-pollination syndromes (Burd and Allen 1988), plant geometry (Bickel and Freeman 1993) or variation in environmental quality (Guillon et al. 2006). However, even if pollination biology is known to strongly shape floral characters (Raven et al. 2005), no studies have tested yet its influence on the application of the SAM within the context of biotic pollen vectors. This gap

might be due to the scarcity of studies combining careful examination of pollinators with accurate surveys of flower allometric characters, but also to the complexity of some mathematical models developed to understand the SAM (Guillon et al. 2006).

Despite the apparent complexity of this topic, some simple predictions can be made regarding the possible influence of both the efficiency and diversity of pollinators on the investment towards a given gamete type. Lloyd and Bawa (1984) first proposed that a gamete positive-bias is expected for the gamete dispersing longer distances, since this would avoid intra-plant gamete competition. Here, we re-visit this hypothesis from the point of view of the pollinators. In an exclusive zoophilous plant, in which pollen is strictly dispersed by animals, pollen dispersion is directly related to pollinator efficiency (a function of both the number of pollinators visiting a flower and of their individual capability to effectively cross-pollinate). As a consequence, if the probability of cross-pollination is low because pollinators are scarce, present a low ability to cross-pollinate, or both, we predict that the SAM should apply: gametes should be biased towards female function because pollen dispersion is expected to be low. In addition, by investing on female flowers, the number of pollinated ovules should be quickly increased as soon as an eventual pollen-carrier visits the flower. Furthermore, an investment towards male gametes in such a situation would translate into a large loss of these gametes. On the other hand, if the probability of cross-pollination is high because pollinators are very abundant, cross-pollination is efficient, or both, plants should not match to the SAM, since it is likely that the pollen produced will present a high probability of reaching another flower, diminishing a potential intra-plant gamete competition. Conforming to the SAM in such a context is expected to be counter-selected because the plant has the possibility to favour both its fruit set and male gamete dispersion (Chartier and Gibernau 2009).

As stated above and in addition to pollinator efficiency, we also consider pollinator diversity (in terms of number of different taxa visiting a flower) as a possible factor influencing the application of the SAM. In fact, the overall probability of cross-pollination is explained by the interaction of two main factors: pollinator diversity and efficiency. Pollinator diversity in a given plant can be considered as a proxy for the level of specificity in a pollination interaction (along a continuum between generalist and species-specific pollination, Ollerton et al. 2007), and thus indirectly informs on the probability that a given pollinator will carry specific pollen. The nature of this interaction (*i.e.*, specific or generalist) combined with the individual efficiency of the pollinator (*i.e.*, its visitation rate and constancy) finally shapes the probability of cross-pollination for a given pollinator. We expect that very efficient and specific pollinators would provide a final global high probability of cross-pollination, which would not favour the application of the SAM (high pollen dispersion, low intra-plant gamete competition), whereas selection towards SAM expression should be advantaged under generalist and inefficient visits (low pollen dispersion and high intra-plant gamete competition). Moreover, the shape of the probability curve for the application of the SAM is expected to be different for cases in which specific *vs.* generalist pollinators are involved (i.e., with a faster decrease of intra-plant pollen competition in the case of specific visitors): if pollen load gets saturated the visit of a reduced number of pollinators from a few species will be translated into a higher relative probability of receiving pollen from the “correct” species than if visitors were generalists. Predictions regarding the application of the SAM as a function of both pollinator diversity and efficiency are summarized in Figure 1.

Geophytic species of the Araceae are well-suited to test hypotheses within the SAM framework since inflorescence and plant-size in this plant family are tightly related to reserves stored in the tuber (Méndez and Obeso 1993; Vitt et al. 2003)

and increase with the age of the plant. Related to this point, it has been demonstrated for example that *Arum italicum* Mill. and *A. idaeum* Coust. & Gand. do not produce any inflorescence until the tuber mass has reached a threshold weight, after which larger plants produce more inflorescences than smaller ones (Méndez and Obeso 1993). This physiological feature would indicate that there are genetic constraints intrinsic to the species which make the relationship bulb size - reproductive allocation specific to a species and less variable at a regional level.

The simplest view of the SAM in the case of Araceae would thus predict a trend towards feminization in larger inflorescences. Within bulbous Araceae, the genus *Arum* is particularly suitable for testing hypotheses related to the effect of pollinator efficiency and diversity on the SAM. It includes protogynous, thus generally self-infertile species, which present similar ecologies and biologies (Boyce 1993) and express insect-pollination syndromes with different degrees of pollinator specificity (Gibernau et al. 2004). However, the application of the SAM is heterogeneous within the genus (see below). The genus *Arum* comprises 28 species, most of them with a peri-Mediterranean distribution (Boyce 1993; Espíndola et al. 2010). As all members of the Araceae family, plants are characterized by singular floral structures, including a spadix harbouring flowers, surrounded by a spathe, *i.e.*, a modified bract (Mayo et al. 1997). A special feature of the genus is that this basic floral structure is adapted to the attraction and trapping of pollinators (lure-and-trap antagonistic pollination). In all but one species (*i.e.*, *A. creticum* Boiss. & Heldr.) a contraction of the spathe takes place just above the fertile flowers, where sterile flowers transformed into hairs (*i.e.*, staminodes) delimitate a closed floral chamber (Boyce 1993). Another remarkable feature of *Arum* species is their ability to produce heat and odour during anthesis in order to attract and temporarily capture potential pollinators (Gibernau et al. 2004). Probably because of the substantial amount of energy involved both in the

158 characteristic floral structure and in the heat and odour production, a trend towards
159 specificity in pollinator attraction is generally detected (Gibernau 2003). However,
160 there is strong variation in pollinator diversity levels across the different *Arum*
161 species. At one end of the continuum are species such as the specialist *A.*
162 *maculatum* L., a widespread European species, which shows a highly specific
163 interaction with two species of the tribe Psychodini [*i.e.*, *Psychoda phalaenoides* L.
164 and *Psycha grisescens* Tonnoir (Diptera: Psychodidae); Espíndola et al. in press]. At
165 the other end are species such as the generalist *A. italicum* Mill. –also widely
166 distributed in Europe– whose efficiency in attracting pollinators is lower and is visited
167 by several arthropods from different families and orders (Albre et al. 2003). Whereas
168 the latter species has been shown to follow the SAM (Albre and Gibernau 2008;
169 Méndez 2001), the first does not do so (Chartier and Gibernau 2009). These two
170 species therefore confirm predictions illustrated in Figure 1: plants visited by efficient
171 and specific pollinators should not favour the SAM, while those visited by rare and
172 generalist pollinators should increase their fitness by following this model. However,
173 it is difficult to confirm this theory based on these two single cases. For this reason,
174 we propose to further test our predictions by investigating another widely-distributed
175 European *Arum* species, *A. cylindraceum* Gasp. The choice of this species is based
176 on its wide distribution range similar to those of the other currently well-known *Arum*
177 species, thus decreasing possible effects of other untested external factors (*e.g.*,
178 environmental features, local adaptations) on the application of the SAM.

179 *Arum cylindraceum* is a recently rediscovered and redescribed species (Bedalov
180 and Küpfer 2005; Fridlender 1999), whose ecology and biology are currently not well
181 known. Its centre of distribution is Eastern Europe, being present from Spain and the
182 Balkans to Poland and Denmark (Bedalov and Küpfer 2005). Until now, no studies
183 have addressed the question of the identity of pollinators in this species, even if it

likely presents some level of specificity, as shown in other species within *Arum* (Gibernau et al. 2004). In the present work, we investigate the pollination strategy of *A. cylindraceum* at the species level and use it as a third case study to test the validity and wider application of our predictions on the SAM as a function of pollinator efficiency and diversity. In order to identify an eventual specificity in the pollination interaction and to get a large-scale overview of its geographical consistency, we studied the species composition in arthropod visitors –and thus potential pollinators– throughout the distribution range of the plant. This allowed us evaluating both the reproductive strategy of the species (*i.e.*, specific *vs.* generalist) and the level of efficiency in pollination as a function of the number of visitors attracted. Using these cues and applying the predictions presented above (see Fig. 1) we compared results in efficiency and diversity of pollinators with allometric measures taken on the inflorescences of the sampled plants and tested the following expectations: i) a combination of high insect diversity and low number of visitors is expected to be found in plants conforming to the SAM, because the probability of cross-pollination is expected to be low (as in *A. italicum*); ii) a low insect diversity combined with a high number of visitors should be related with the non-conformity to the SAM, since the probability of cross-pollination is expected to be high (like in *A. maculatum*); iii) a combination of low insect diversity and number is expected in plants matching to the SAM (low net probability of cross-pollination, especially if the number of pollinators is extremely low); iv) a combination of high insect diversity and number is expected to be present in plants which do not match to the SAM (high net probabilities of cross-pollination).

Material and methods

Sampling and pollinator identification

Arum cylindraceum populations were sampled from April to May 2008, almost completely covering its distribution range. Considering the generally small population sizes (less than 80 plants) and highly aggregated distribution of the plants in populations, we excluded the effect that clonal reproduction (a feature known for the genus, Boyce 1993) could have in our results by restricting the number and distribution of sampled individuals. This is why we selected up to five floral chambers per population from plants distant by more than five meters from their closest sampled neighbour. Because we studied the complete distribution range of the plant, and since we were interested in having a view of the strategy displayed by the species in general and not necessarily at each sampled location, we consider that this sampling is representative of the main strategy displayed by the species and has the advantage of not being biased by locally high clonal reproduction. Once floral chambers were collected, they were labelled and independently preserved in a Falcon tube filled with 70% ethanol. A plant voucher was collected at each location, and geographic coordinates were registered using a GPS Etrex Garmin.

Arthropods were identified to the order level (Freude et al. 1964). Diptera families were indentified at the species level. Because of the lack of updated identification keys for several Diptera families, European specialists were consulted for the identification of samples belonging to this group. An exception is the Psychodidae family, for which keys are actually available (Jezek 1990; Withers 1988). Occurrences of each species/genus/family were recorded and numbers per floral chamber, per population and in total were calculated, as well as the sex-ratio of the most abundant visitors.

Because the identity and amount of pollinators have been shown to vary geographically in other *Arum* species (Espíndola et al. in press) we tested for an

eventual spatial structure by analyzing the geographic visitor composition using ArcMap 9.3 (ESRI, Redlands CA, USA).

Allometric measures

Out of all analyzed plants, a maximum of three inflorescences per population (n=39) were selected to perform allometric measures (lengths of the spadix -see below-, female and male zones). Numbers of male and female flowers were counted and floral sex ratios calculated. The floral sex ratio was inferred as: number of male flowers/(number of male flowers + number of female flowers), where 1 represents a complete male, and 0 indicates a complete female. Following the approach established by Chartier and Gibernau (2009), regions of the inflorescence were divided into fertile male and female zones, and their lengths were measured with a digital calliper (TESA, model CCMA-M). Because the addition of fertile and sterile zones is correlated in a linear manner with the total size of the inflorescence (M. Gibernau, unpublished), we measured the length from the base of the female fertile zone to the last upper staminodes and used this value as a representative of the total spadix length, a value correlated with the total plant size, and thus with available resources (Albre et al. 2003).

Descriptive statistics of these measures and correlations (linear regressions for lengths; logistic regressions for counts) were calculated, applying the functions lm and glm implemented in R 2.9.0 (R Core Development Team, 2009). Because the present study was interested on general floral patterns intrinsic to the species, and because values presented no special population differentiation (see below), all analyses were done at the interpopulation level.

Testing predictions

In order to test predictions made by our model, we applied a comparative method. Based on the data obtained from the insect identification, we inferred their

diversity and efficiency. Diversity was inferred following the approach applied by Ollerton et al. (2007), which defined the specificity as a measure of the number of interacting species. Efficiency was estimated by considering the total number, and the constancy in the composition of visitors from a spatial point of view, since it is unlikely that a group of rare, diverse and inconstant visitors presents a high overall pollination efficiency, while the opposite is true for abundant and specific visitors.

The correctness of the predictive model (Fig. 1) was thus inferred by using indirect measures of diversity and efficiency to identify the theoretical expectations corresponding to the different combinations addressed, and further comparing them with the results obtained in the allometric survey. Congruence between these allometric results and expectations would provide support to the model.

Results

Visitors' diversity and distribution

A total of 24 populations were sampled (Appendix 1) and 80 floral chambers collected, containing 252 arthropods distributed in half of the populations (12 populations did not comprise any visitor). This represented a mean of 3.50 (SE=0.57) and 12.75 (SE= 3.64) arthropods per floral chamber and per population, respectively.

Figure 2 shows the diversity in collected visitors, restricted to Arthropods and represented by the classes Insecta (67.9%), Arachnida (Acarina, 18.2%), Diplura (7.1%) and Collembola (6.7%). The two most frequent orders were Diptera (65.5%) and Coleoptera (28.1%). The order Diptera comprised families Psychodidae (57.1%), Ceratopogonidae (14.3%), Sphaeroceridae (10.7%), Sciaridae (7.1%), Cecidomyiidae (3.6%), Chironomidae (3.6%), and Mycetophilidae and Phoridae (3.6%).

Regarding the diversity found in each Diptera family, our results show that Psychodids were represented by a large majority of *Psycha grisescens* (82.8%, mainly males) and lower proportions of females of *Psychoda phalaenoides* (14.1%) and *P. albipennis* Zett. (3.1%). Ceratopogonids were identified as belonging to the species *Culicoides obsoletus* Meigen (94%) and *Atrichopogon oedemerarum* Stora (6%). Species present in the family Sphaeroceridae were *Spelobia clunipes* Meigen (58.3%), *Coproica ferruginata* Stenhammar (33.3%) and *C. vagans* Haliday (8.4%). Concerning the Sciarids present in the samples, we could only identify them to the two genera *Bradysia* Winnertz (63%) and *Phytosciara* Frey (25%). Samples of the remaining families were very scarce (less than 11% in total): taxa were *Smittia leucopogon* Meigen (Chironomidae), species belonging to genera *Azana* Walker and *Boletina* Staeger. (Mycetophilidae) and to tribes Orthocladiinae (Chironomidae), Lestremiinae and Cecidomyiinae (Cecidomyiidae). No sub-generic identification was possible on the unique Phorid sample.

The spatial distribution of these groups is not geographically structured in terms of composition (Fig. 3a) or of number of visitors trapped (Fig. 3b).

Allometric measures

There is a logical positive correlation between the length of each floral zone (*i.e.*, male, female and sterile) and the total inflorescence length (Fig. 4a). The increase in the length of the sterile floral zone was however the highest (slope=0.562), followed by female (slope=0.334), and finally by male (slope=0.103) zones. Each factor significantly increased the adjustment of the general linear model indicating that the three slopes are different ($p<0.001$ in all the cases, $R^2=0.946$, $F=379.2$).

Another positive relation is observed for the number of fertile flowers and the total inflorescence length (Fig. 4b). The slope is higher for the number of female

(slope= 0.462) than for male (slope=0.225) flowers, and both slopes are significantly different ($p<0.05$, $z_{(fem)}= 5.781$, $z_{(male)}= -2.425$).

The inflorescence floral sex-ratio is negatively correlated (slope= -0.063, $R^2=0.094$, $F=4.626$, $p<0.05$) with the total inflorescence length (Fig. 4c), indicating a significant feminization of the inflorescence with an increase of the total size.

There was no significant correlation between the number of trapped insects and the total length of the inflorescence ($p>0.1$; Fig. 4d). This result was maintained even when samples presenting or not insects were considered (not shown) and despite the fact that mean total lengths were significantly different between the two groups ($F=4.992$; $df=1$; $p<0.05$).

Correlations showed no population structure (data not shown), indicating that the observed patterns are intrinsic to the species and not to a specific region.

Interaction between SAM and visitors

In a comparative framework, our results showed that *A. cylindraceum* presents a visitor pattern combining the presence of high visitor diversities (Fig. 2) with low numbers of spatially highly variable compositions of visitors (low expected efficiency, Fig. 3), which represents the combination “high diversity – low efficiency” predicting the application of the SAM (Fig. 1).

The survey on the variation of reproductive allocation indicates that the plant modifies the proportion of one or the other flower type in function of the available resources (Figure 4a-c), indicating that the SAM is applied in *A. cylindraceum*.

Discussion

The diversity and efficiency of *A. cylindraceum* visitors

Compared to other *Arum* species (Espíndola et al. in press; Albre et al., 2003), our results show that visitors of *A. cylindraceum* are scarce (3.5 per floral chamber and completely absent of half of the populations) and highly diverse comprising species from several insect groups. Because the plant does not seem prone to self-fertilization in the European continent (Chouteau et al. 2008; Fridlender 1999), such a low number of trapped insects was not expected. Since several species within the genus can reproduce by forming bulbs in a vegetative manner (Boyce 1993), it is likely that *A. cylindraceum* can reproduce by clonal means (Fridlender 1999) and that cross-pollination may rarely happen. However, during our sampling trips, it was frequent to find *A. cylindraceum* plants displaying fruits (A. Espíndola, pers. obs.). Another explanation for such a low number of sampled insects could be that this species is effectively visited by more arthropods than those found in the present study, but that the trapping system (*i.e.*, the number and length of staminodes) is not highly efficient, releasing arthropods after a short span of time spent in the floral chamber. Such a non-optimal strategy would require a smaller energetic investment in terms of odour (*e.g.* less specific volatiles), heat and morphological features (*e.g.* length of staminodes). It also might be sufficient to ensure pollination of several female flowers during the lifespan of a plant, if at least one insect carrying pollen visits the inflorescence (i.e., as shown in *A. maculatum* (Lack and Diaz 1991)). However, this strategy would bring the problem that if visitors are not trapped quickly after the first trapping, pollen would lose its viability (Gibernau and Macquart 2003) and pollination would not occur if the pollen carrier visited a second flower. Long-lasting observation and sampling of all insects visiting the flowers during the whole anthesis would allow testing these ideas. Additionally, investigating the variation of pollen viability through time would also inform on the temporal range during which

pollen carry-over needs to happen, and would provide indirect information on the natural dynamics of pollination.

Despite showing a low total number of visitors, our results still reveal a high diversity of insects and especially of Diptera (Fig. 2), as shown for some other *Arum* species (Gibernau et al. 2004). All Dipteran families recorded here (with the exception of the three less abundant ones, *i.e.*, Cecidomyiidae, Mycetophilidae and Phoridae) have been previously identified as visitors and thus potential pollinators of several other species within the genus, as for example *A. dioscoridis* and *A. orientale* (for a review, see Gibernau et al. 2004). Moreover, their simultaneous occurrence was also recorded in the floral chambers of *A. italicum*, where they were confirmed as effective pollinators (Albre et al. 2003). Considering these reasons, we feel confident to consider that insects trapped in the present study are very probably reliable pollinators of *A. cylindraceum*.

In contrast to a recent study on *A. maculatum* (Espíndola et al. in press), the spatial distribution of visitors' composition does not show any specific geographic pattern (Fig. 3), indicating that the pollination strategy applied by the plant is intrinsic to the species and is not a result of local adaptation. Consequently, the plant appears as a widespread generalist, presenting little efficiency in pollinator attraction.

Arum cylindraceum **in the SAM perspective**

Because of its low efficiency and generalist behaviour, *A. cylindraceum* should be expected to conform to the SAM (Fig. 1). This prediction is confirmed by our results on the variation in the resource allocation applied by the plant (Fig. 4): there is an increasing investment towards female function as inflorescence size becomes larger. As shown in Fig. 4b, the increase in the female floral zone is mostly due to an increase in the number of female flowers, indicating that not only the size but also the number of female flowers increases in larger plants. In agreement with the

389 similar pattern found in *A. italicum* (Albre and Gibernau 2008), the mean floral sex
390 ratio of the plant is biased towards the male function, but significantly feminizes
391 when the inflorescence size increases (Fig. 4c). This trend holds even if both flower
392 types increase in number (Fig. 4b). Our simple SAM-pollinator model is again
393 confirmed, indicating that its rules seem biologically correct.

394 In *Arum*, it has been demonstrated that the augmentation in plant size is
395 associated with an increase in the size of the organ involved in heat and odour
396 production -the appendix- (Gibernau and Albre 2008), and a subsequent raise of
397 pollinator attraction (Méndez and Obeso 1992). Our analyses however show that in
398 *A. cylindraceum* the increase in inflorescence size is not correlated with a gain in
399 attraction power (Fig. 4d). Moreover, this result corroborates our view of the
400 interaction between *A. cylindraceum* and its associated pollinators as a low-
401 efficiency pollination system.

402 In the same way that we could not identify any geographic structure in the
403 composition of insects, we could not find any local variation of floral characters
404 (samples from the same locations were randomly distributed in the point clouds; Fig.
405 4). Moreover, points from all locations were clearly distributed along a unique and
406 continuous regression line, which would not be expected if there were local
407 specializations. This would indicate –as previously thought- that the relationship
408 between bulb size and floral allocation is genetic and thus intrinsic to the species. A
409 confirmation of this relationship would however require measuring bulb weights and
410 floral allocations both at local and general scales.

411 The combination of pollinator composition and plant allometric measures
412 confirmed our expectations in long-lived zoophilous species. This is to our
413 knowledge the first time that this theory is both explicitly proposed and tested in an
414 empirical pollination system. It is important to note that our predictions in Araceae

might also apply to other plant groups, as for example to species of the genus *Trillium* L. (Melanthiaceae) (Wright and Barrett 1999), in which low pollination visitation with application of the SAM has been registered. Nevertheless, in order to extend the understanding of the dynamics of the SAM as a function of pollinator diversity and efficiency, it is necessary to analyze a wider number of plant-insect systems taking into account not only the variation of sex-resource allocation, but also pollinator diversity. In this context, the family Araceae is particularly suitable because biological (e.g., long-lived plants) and ecological characters (e.g., generalised zoophily) shared by its species greatly simplify the comprehension of SAM-related floral allocation responses. Finally, the application and development of dynamic mathematical models in the SAM/pollinator context could provide important information on the exact expected shapes of the probability curves proposed by our predictions, and would thus help deepen our understanding of conditions guiding the variation of floral allocation.

Acknowledgements

The authors are thankful to Dr. Kącki, Dr. Dajdok (University of Wrocław, Poland), Dr. Triponez and Prof. Küpfer (University of Neuchâtel, Switzerland) for their invaluable technical assistance during field work. The present study was funded by the Swiss National Science Foundation (project No. 3100A0-116778), the Matthey-Wütrich Grant of the University of Neuchâtel accorded to N.R. and the Travel Grant of the Swiss Academy of Sciences SCNAT+ accorded to A.E.

438 **References**

439 Albre J, Gibernau M (2008) Reproductive biology of *Arum italicum* (Araceae) in the
440 south of France. Bot J Linn Soc 156:43–49
441 Albre J, Quilichini A, Gibernau M (2003) Pollination ecology of *Arum italicum*
442 (Araceae). Bot J Linn Soc 141:205–214
443 Bedalov M, Küpfer P (2005) Studies on the genus *Arum* (Araceae). Bull Soc Neu Sci
444 Nat 128:43–70
445 Bickel AM, Freeman DC (1993) Effects of pollen vector and plant geometry on floral
446 sex ratio in monoecious plants. Am Midl Nat 130:239–247
447 Boyce PC (1993) The genus *Arum*. The Royal Botanical Gardens, Kew, London, UK.
448 Burd M, Allen TFH (1988) Sexual allocation strategy in wind-pollinated plants.
449 Evolution 42:403–407
450 Chartier M, Gibernau M (2009) Size variations of flowering characters in *Arum*
451 *maculatum* (Araceae). Aroideana 32:153–158
452 Chouteau M, Gibernau M, Barabé D (2008) Relationships between floral characters,
453 pollination mechanisms, life forms, and habitats in Araceae. Bot J Linn Soc 156:29–
454 42
455 Espíndola A, Buerki S, Küpfer P et al. (2010) New insights into the phylogeny and
456 biogeography of *Arum* L. (Araceae): unravelling its evolutionary history. Bot J Linn
457 Soc 163:14–32
458 Espíndola A, Pellissier L, Alvarez, N. Variation in the proportion of flower visitors of
459 *Arum maculatum* along its distributional range in relation with community-based
460 climatic niche analyses. OIKOS (in press)
461 Freude H, Harde KW, Lohse GA (1964) Die Käfer Mitteleuropas. Goecke & Evers,
462 Krefeld.

463 Fridlender A (1999) Observations sur la biologie de l'*Arum cylindraceum* (Gasp.)
464 Araceae en Corse. Acta Bot Gallica 146:297–309
465 Ghiselin MT (1969) The evolution of hermaphroditism among animals. Q Rev Biol
466 44:189–208
467 Gibernau M (2003) Pollinators and visitors of Aroid inflorescences. Aroideana 26:73–
468 91
469 Gibernau M, Albre J (2008) Size variations of flowering characters in *Arum italicum*
470 (Araceae). Aroideana 31:101–106
471 Gibernau M, Macquart D (2003) Pollen viability and longevity in two species of *Arum*.
472 Aroideana:58–62
473 Gibernau M, Macquart D, Przetak G (2004) Pollination in the genus *Arum* - a review.
474 Aroideana 27:148–166
475 Guillon JM, Julliard R, Leturque H (2006) Evolution of habitat-dependent sex
476 allocation in plants: superficially similar to, but intrinsically different from animals. J
477 Evol Biol 19:500–512
478 Jezek J (1990) Redescriptions of nine common palaearctic and holarctic species of
479 Psychodini End. (Diptera: Psychodidae). Acta Entom Mus Nat Pr 43:33–83
480 Lack AJ, Diaz A (1991) The pollination of *Arum maculatum* L. - a historical review
481 and new observations. Watsonia 18:333–342
482 Lloyd DG, Bawa KS (1984) Modification of the gender of seed plants in varying
483 conditions. Evol Biol 17:255–338
484 Mayo SJ, Bogner J, Boyce PC (1997) The genera of Araceae. Royal Botanical
485 Gardens, Kew, London, UK.
486 Méndez M (2001) Sexual mass allocation in species with inflorescences as
487 pollination units: a comparison between *Arum italicum* and *Arisaema* (Araceae). Am
488 J Bot 88:1781–1785

Méndez M, Obeso JR (1992) Influencia del osmóforo en la producción de infructescencias en *Arum italicum* Miller (Araceae). Ann J Bot Mad 50:229–237

Méndez M, Obeso JR (1993) Size-dependent reproductive and vegetative allocation in *Arum italicum* (Araceae). 71:309–314

Ollerton J, Killick A, Lamborn E, et al. (2007) Multiple meanings and modes: on the many ways to be a generalist flower. Taxon 56:1–12

Policansky D (1982) Sex change in plants and animals. Ann Rev Ecol Syst 13:471–495

Primack RB, McCall C (1986) Gender variation in a red maple population (Acer rubrum; Aceraceae): a seven-year study of a "polygamodioecious" species. Am J Bot 73:1239–1248

Raven PH, Evert RF, Eichhorn SE (2005) Biology of plants. W.H. Freeman, New York, NY.

Vitt P, Holsinger KE, Jones CS (2003) Local differentiation and plasticity in size and sex expression in Jack-in-the-pulpit, *Arisaema triphyllum* (Araceae). Am J Bot 90:1729–1735

Warner RR (1988) Sex change in fishes: hypotheses, evidence, and objections. Environ Biol Fish 22:81–91

Withers P (1988) Moth flies. Diptera: Psychodidae. Dipt Dig 4:1–83

Wright SI, Barrett SCH (1999) Size-dependent gender modification in a hermaphroditic perennial herb. Proc R Soc Lond B 266:225–232

511 Appendix 1 – Sampled locations.

Locality	Country	Collection date	Altitude (m)	Latitude	Longitude
Mont Iti Kato	Greece	20.07.2007	1377	38.80973	22.20632
Gross Disnack	Germany	15.04.2008	42	53.72313	10.70769
Schmilau	Germany	15.04.2008	-8	53.66833	10.7575
Purkersdorf	Austria	16.04.2008	283	48.20223	16.17298
Herl'any	Slovakia	20.04.2008	300	48.79885	21.48015
Tylawa	Poland	22.04.2008	350	49.46503	21.7323
Sisteron	France	26.04.2008	508	44.21825	5.91398
Vondrou	Greece	25.05.2008	685	41.27574	23.74667
Chaliki	Greece	26.05.2008	1132	39.68423	21.19218
Delvinaki	Greece	29.06.2008	827	39.92821	20.46182
Anilio	Greece	26.05.2008	1369	39.73474	21.19969
Sokobanja Vrh	Serbia	23.05.2008	917	43.58153	21.89266
Kalvehale	Denmark	24.04.2008	2	54.99954	12.16737
Villetta Barrea	Italy	28.05.2008	1258	41.79683	13.93833
Gargano	Italy	29.05.2007	703	41.7937	15.97829
Zobor	Slovakia	19.04.2008	310	48.33696	18.09889
Gioia Vecchio	Italy	29.05.2008	1390	41.90208	13.73339
Bucje	Serbia	22.05.2008	669	44.15512	22.11236
Gylling	Denmark	25.04.2008	25	55.84845	10.16711
Traiskirchen	Austria	17.04.2008	207	48.00996	16.29883
Solosnica	Slovakia	18.04.2008	170	48.4538	17.23844
Muszkowice	Poland	23.04.2008	226	50.64387	16.95121
Church of Taxiarches	Greece	30.04.2007	832	32.87914	20.67902
Trpejca	Macedonia	25.05.2008	1361	40.96949	20.85911

512

513 Table 1 – Descriptive statistics of allometric measures in inflorescences of *Arum*

514 *cylindraceum*. SE: Standard Error.

		Mean	SE
Number of Flowers	**female**	38.92	1.67
	male	71.61	2.70
Floral zone length (cm)	**female**	0.92	0.028
	male	0.46	0.012
	sterile	1.44	0.039
Total length (cm)[1]		2.81	0.035
Sex-Ratio		0.65	0.011

515 [1] from the base of the female zone up to the highest staminode (not including the

516 stipe and the appendix).

517

518 Figures legends

519 Figure 1 – Predictions of application of the size-advantage model, considering
520 pollinator diversity and efficiency in zoophilous plants. Because variables are
521 continuous, grey scales indicate the expected variation of the probability of
522 application of the model (dark: high probability, white: low probability). Note that in
523 specific interactions, the application of the SAM is expected to be advantageous
524 mainly at very low visitation rates (low net pollination efficiency), since specificity in
525 plant-insect interactions is generally associated with highly efficient individual cross-
526 pollination due to coevolutionary processes.

527

528 Figure 2 – Visitor composition of *Arum cylindraceum* floral chambers in 12
529 populations at three taxonomic levels: Arthropod classes, Insecta orders, Diptera
530 families. Total numbers of specimens are indicated below each group.

531

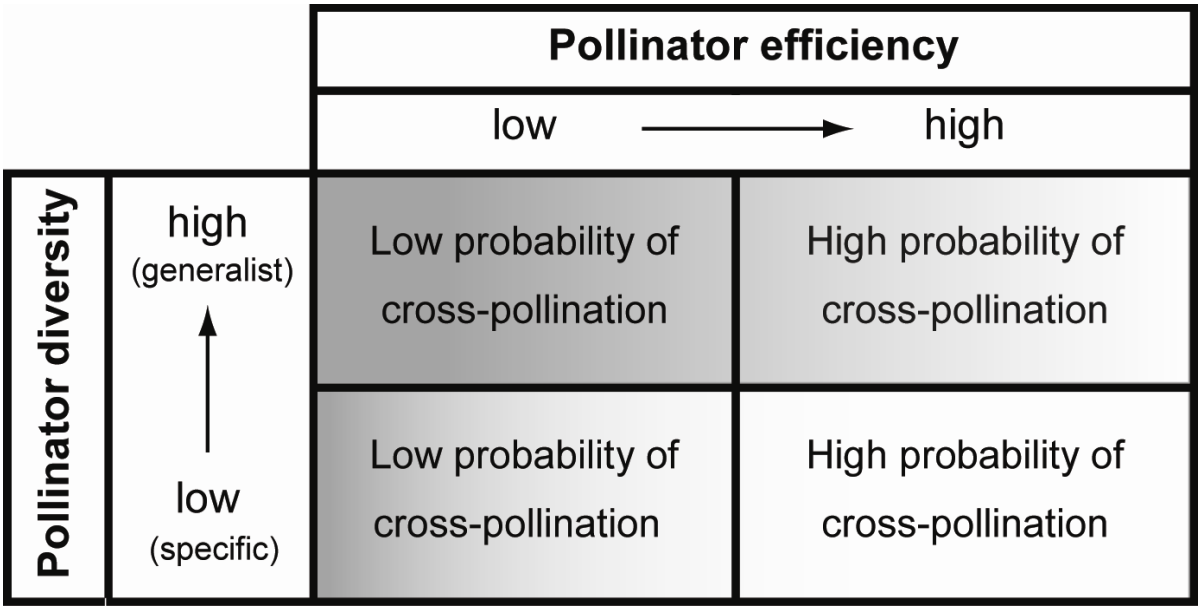
532 Figure 3 – Visitor composition in the sampled populations. a) among Arthropod
533 orders; b) among Dipteran families. Pie sizes in b are proportional to numbers of
534 trapped insects (the scale is referred in the legend).

535

536 Figure 4 – Allometric measures of *A. cylindraceum* inflorescences. a. Length of
537 each floral region (male, female, sterile) in function of the total inflorescence length:
538 male (filled black circles; T_{slope}=7.225, slope=0.103, p<0.001); female (empty circles;
539 T_{slope}=3.695, slope=0.334, p<0.001); sterile (filled grey circles; T_{slope}=10.722,
540 slope=0.562, p<0.001). b. Number of female (empty circles, Z=5.781, slope=0.462,
541 p<0.001) and male (filled circles, Z=-2.425, slope=0.225, p<0.05) flowers in function
542 of total inflorescence length. c. Floral ratio in function of total inflorescence length
543 (T=-2.151, slope=-0.063, p<0.05). d. Number of insects trapped in function of total

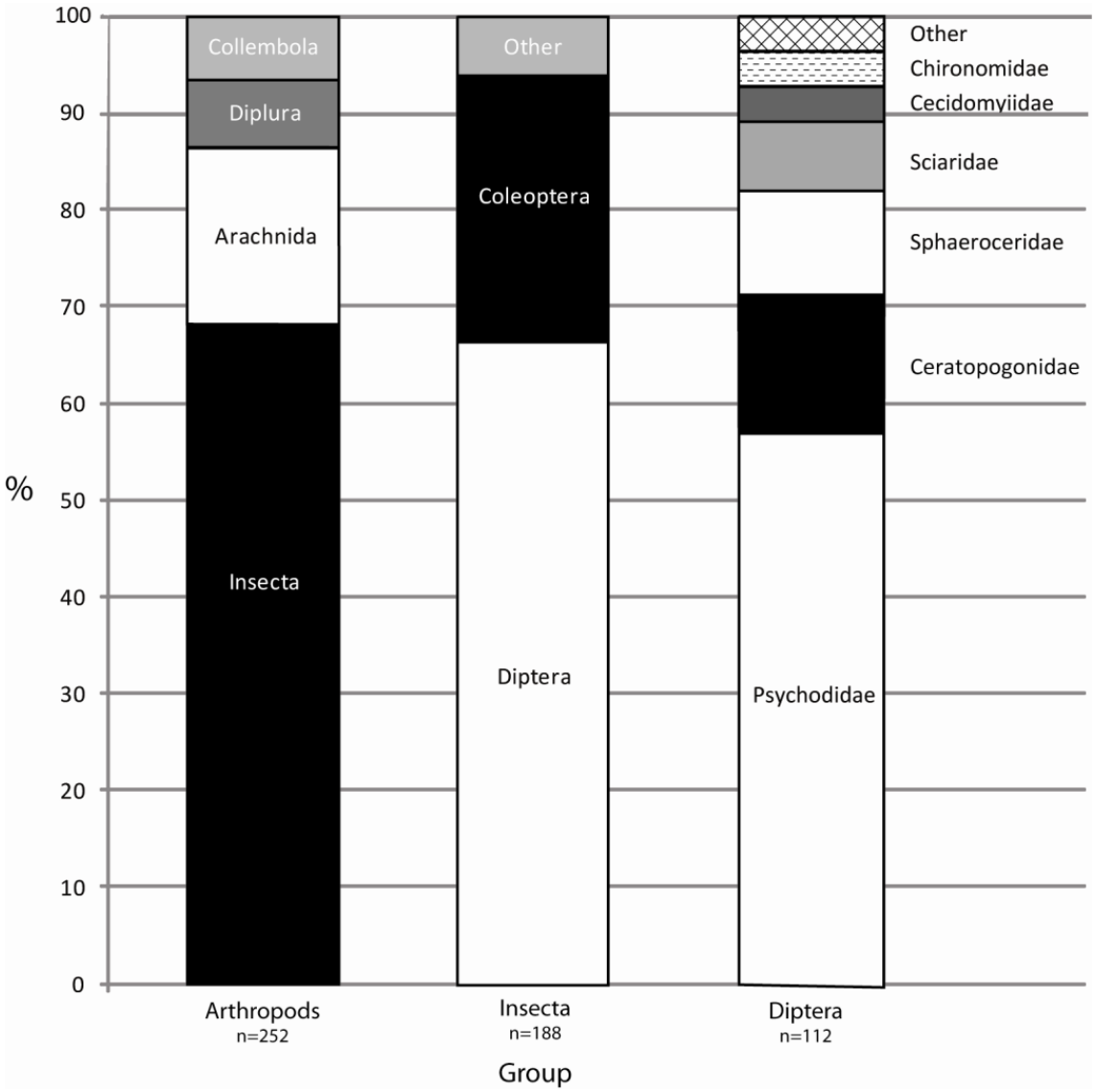
544 inflorescence length (Z=2.819, $p>0.05$; filled circles: inflorescences containing
545 visitors; empty circles: inflorescences not containing visitors). The total inflorescence
546 length corresponds from the base of the female zone up to the highest staminode
547 (not including the stipe and the appendix).
548

549


Pollinator efficiency
low
high
Pollinator diversity
high
(generalist)
low
(specific)
Low probability of cross-pollination
High probability of cross-pollination
Low probability of cross-pollination
High probability of cross-pollination

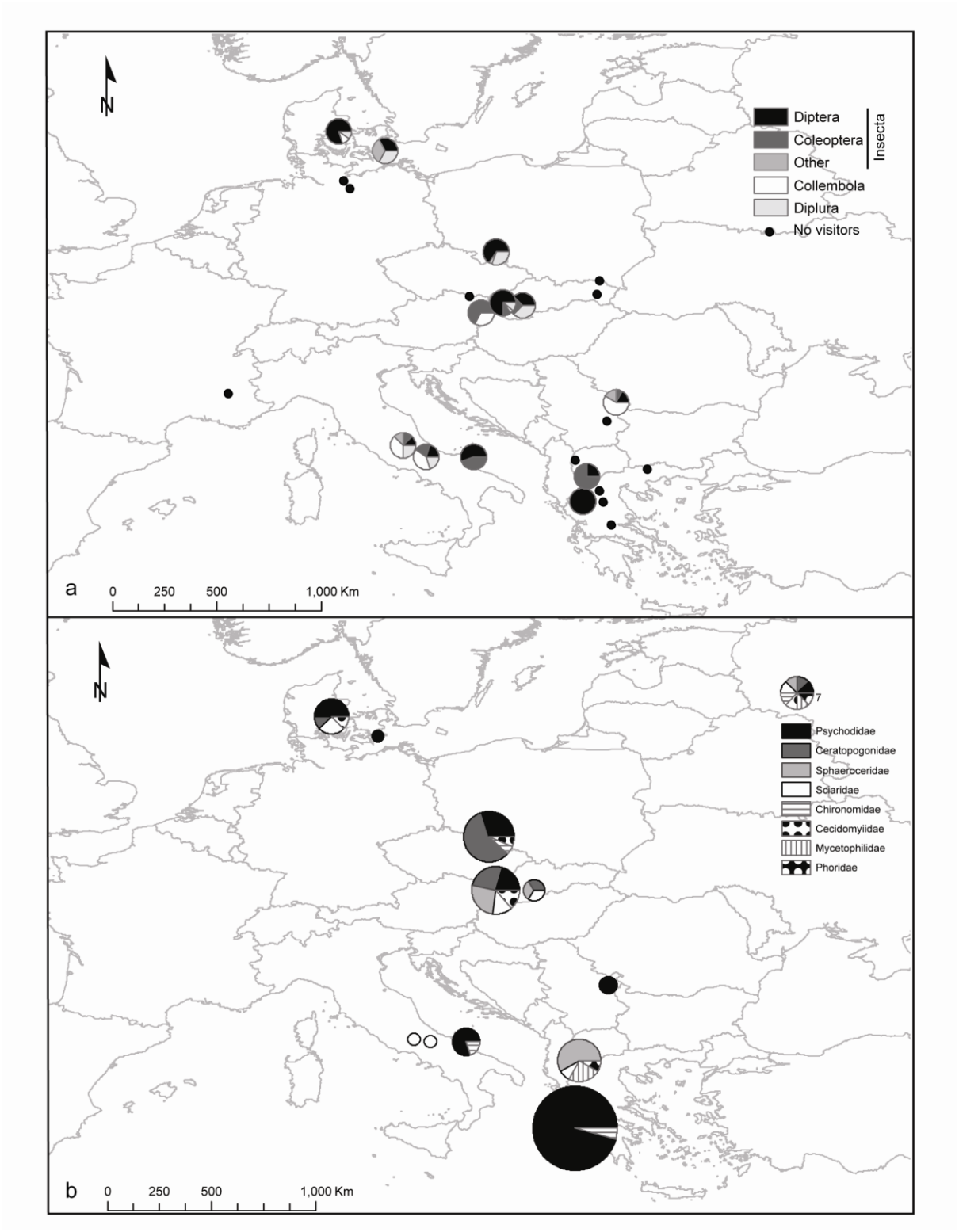
550

551


100
90
80
70
60
50
40
30
20
10
0
%
Collembola
Diplura
Arachnida
Insecta
Other
Coleoptera
Diptera
Other
Chironomidae
Cecidomyiidae
Sciaridae
Sphaeroceridae
Ceratopogonidae
Psychodidae
Arthropods
n=252
Insecta
n=188
Diptera
n=112
Group

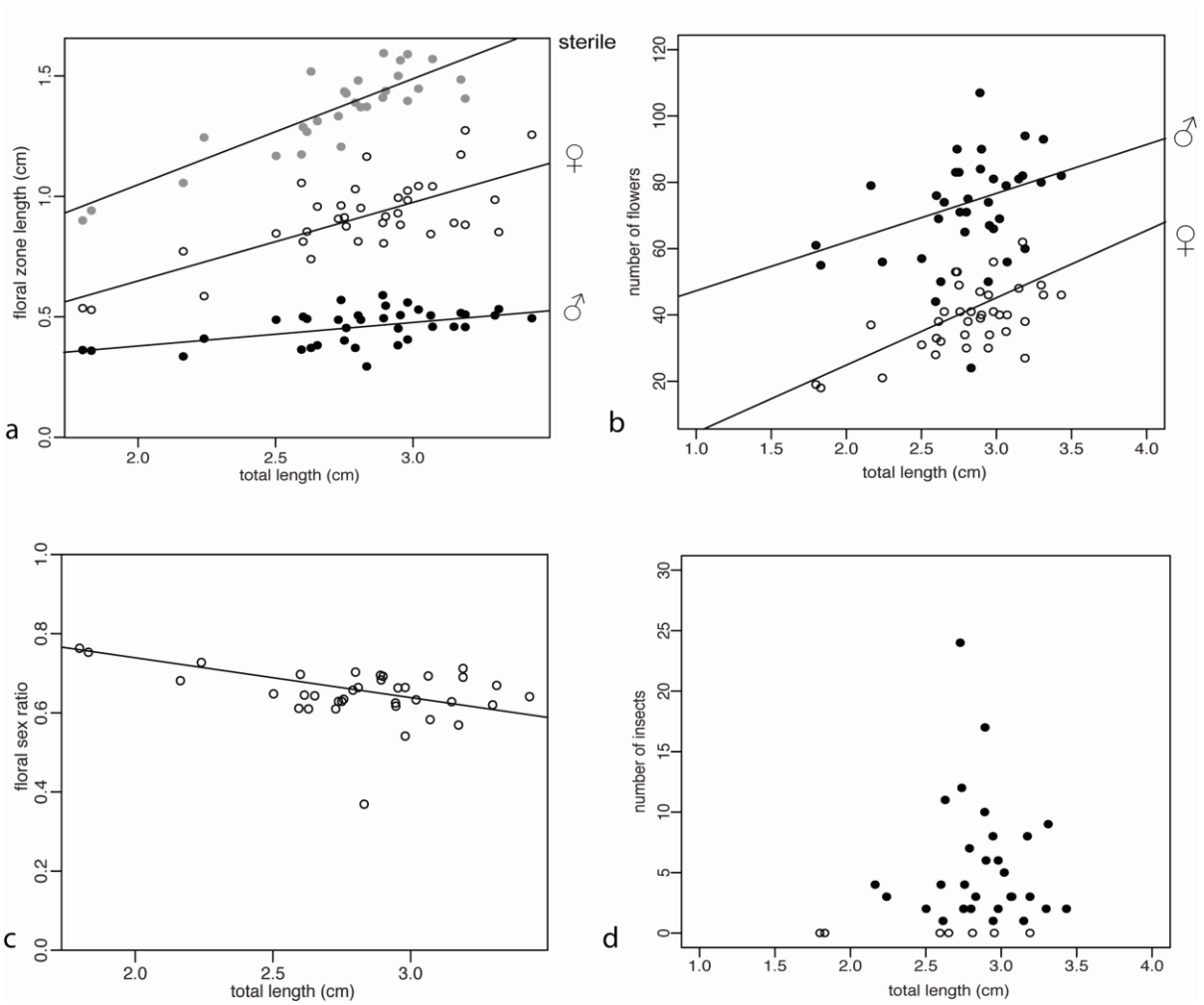
552

553


N
Diptera
Coleoptera
Other
Collembola
Diplura
No visitors
Insecta
a
0
250
500
1,000 Km
N
7
Psychodidae
Ceratopogonidae
Sphaeroceridae
Sciaridae
Chironomidae
Cecidomyiidae
Mycetophilidae
Phoridae
b
0
250
500
1,000 Km

554

555


a
floral zone length (cm)
total length (cm)
sterile
♀
♂
b
number of flowers
total length (cm)
♂
♀
c
floral sex ratio
total length (cm)
d
number of insects
total length (cm)

556

557

Chapter five

Comparative phylogeography in a specific and obligate pollination antagonism

Anahí Espíndola and Nadir Alvarez

Comparative phylogeography in a specific and obligate pollination antagonism

Anahí Espíndola* and Nadir Alvarez†

* Laboratory of Evolutionary Entomology, Institute of Biology, University of Neuchâtel, 2009 Neuchâtel, Switzerland. † Department of Ecology and Evolution, University of Lausanne, Biophore Building, CH-1015 Lausanne, Switzerland.

Keywords: AFLP, *Arum maculatum*, Psychodidae, spatial genetic structure, DNA sequencing, plant-insect interactions

Corresponding author: Anahí Espíndola, Laboratory of Evolutionary Entomology, Institute of Biology, University of Neuchâtel, 2009 Neuchâtel, Switzerland. Fax: +41 (0)32 718 30 01; email :maria.espindola@unine.ch

Running title: Phylogeography of antagonistic interactions

Abstract

The niche of a species is defined not only by climatic and topographic features, but also by biotic components. In specific and obligate pollination interactions, because of the importance of the biotic factor, the presence or absence of a given ecological partner can have important effects on the survival and population dynamics of associated organisms. In a phylogeographic framework, we could thus expect that the fate of organisms interacting specifically is also tightly interrelated. In the present study, we aim at testing such a hypothesis by investigating an obligate plant-insect pollination antagonism. For this, we use the European lure-and-trap system featuring *Arum maculatum* L. (Araceae) and its specific Psychodid (Diptera) visitors *Psychoda phalaenoides* L. and *Psycha grisescens* Tonnoir as a case-study. Based on mitochondrial data (in insects) and AFLP genome fingerprinting (in plants) we first inferred the large-scale phylogeographies of each species and afterwards compared their respective patterns.

Our findings indicate that while the plant presents a geographically well-defined genetic structure, this is not the case of the insects, even though the latter harbour an important amount of genetic variation. This suggests that the phylogeographic histories of the plant and its pollinating insects are not congruent, a result that would indicate that insects move freely during distributional shifts, while plants need either to find sites where insects are already present, or change their pollinating strategy.

Introduction

The study of both recent (e.g., D'Andrea *et al.* 2009) and past (e.g., Nogués-Bravo *et al.* 2008) species distributional shifts has shown that in a changing environment organisms follow the cline of their ecological niche. Environmental variations therefore drive changes in distribution ranges. Because a large amount of data is available for both geological remains and climatic records, we yet know that the Quaternary (2.6 – 0 million years ago -Mya) has hosted series of glacial and interglacial periods, caused by orbital and tectonic events (Ehlers & Gibbard 2007), which thus induced recurrent contractions and expansions of species' distribution ranges (Stewart *et al.* 2010). Because such changes directly impact on demographic population parameters, they can indirectly affect the genetic variation of species through processes such as genetic drift that influence the allelic composition at each location (Hewitt 1996). In addition, when contraction processes lead to fragmentation of a species' distribution area, among-population gene flow might be restricted and local selection could trigger lineages divergence (Stewart *et al.* 2010).

Phylogeographic studies aim at understanding these facets of the evolutionary history of species, with the assumption that the spatial structure of taxa can be explained by historical and geospatial (e.g., topological, hydrological) grounds (Avise 2009). In Europe for example, phylogeographic patterns of organisms have been largely studied for the last 20 years, and general postglacial phylogeographic paradigms have been addressed (see reviews in Hewitt 1999; Schmitt 2007; Taberlet *et al.* 1998). In this region, studies concurred in demonstrating that during climate cooling, temperate species found refugia in the Southern mountainous European peninsulas (i.e., Balkans, Italy and/or Iberia; Taberlet *et al.* 1998), and in some cases also in northern areas (e.g., Carpathians; Stewart & Lister 2001). After ice retreat, recolonization took place through several pathways, with contact zones appearing in several central regions as for instance at natural boundaries such as mountain massifs (Taberlet *et al.* 1998).

However, in a synecological context, it should be considered that species occupy ecological niches that are not solely composed of abiotic factors, but which also present

biotic ecological components (Soberón & Peterson 2005). These latter can thus also potentially shape the genetic structure of species impacting their survival, dispersal and dynamics of populations. This effect could be particularly enhanced in the case of specific and obligate interactions because of among-species ecological interdependence.

Nonetheless, only few studies (e.g., Anderson et al. 2004; Tsai & Manos *in press*) have investigated large-scale phylogeographies of species interacting in specific and obligate interactions. Among these, none has examined the comparative phylogeography of plants and arthropods involved in specific antagonistic interactions, even though these are thought to have driven diversification of several groups of angiosperms and insects in the last million years (Farrell 1998; Farrell *et al.* 1992; Futuyma & Agrawal 2009). In this study, we aim at investigating comparative phylogeography in a European plant-insect specific and obligate antagonistic interaction, using the lure-and-trap pollination system in which the Araceae *Arum maculatum* L. is specifically pollinated by the Psychodid (Diptera) sister species *Psychoda phalaenoides* L. and *Psycha grisescens* Tonnoir (Diaz & Kite 2002; Espíndola *et al. in press*)

In this Aroid, the volatile and heat production, and the presence of special morphological floral structures (Boyce 1993) –a spadix and a spathe modified to create a closed floral chamber, combined with a protogynous maturing dynamics– allows the plant to attract ready-to-lay female Psychodid flies and trap them during anthesis (a maximum of one day), at the end of which pollen release happens. Insects are then covered with pollen and set free; some of them will be trapped a second time by another *Arum* flower, leading to cross-pollination. Because a single plant experiences temporal delay in male and female flower maturation, self-pollination is prevented and the plant completely relies on entomophilous cross-pollination for seed production (Lack & Diaz 1991) during its short flowering period (Ollerton & Diaz 1999). In the case of *Arum*, it is evident that the plant provides no rewards, and insects are thus simply exploited.

Because i) a strong knowledge has been acquired on this European antagonistic lure-an-trap specific and obligate interaction; ii) general phylogeographic patterns are currently well

understood at the scale of the European continent (e.g., Hewitt 1999; Taberlet *et al.* 1998); iii) there is a lack in the understanding of the effect of ecological interactions on phylogeographic patterns, we use here the pollination interaction between *A. maculatum* and its associated Psychodids as a case-study to test hypotheses in the framework of comparative phylogeography. Considering the unilateral obligate nature of the interaction (i.e., the insect is ecologically independent from the plant), we hypothesize that while the insect disperses freely, the plant is strongly constrained in its dispersal and long-term refugial survival by the presence of its associated species; we thus expect that plants and insects demonstrate incongruent phylogeographic patterns (*i.e.*, differences in the number and/or distribution of refugia and recolonization pathways). The alternative hypothesis is that despite this unbalanced dependence, these interacting species disperse in parallel, with the plant following the insect footsteps concurring in a high level of overall congruence in phylogeographic patterns (*i.e.*, similarities in the number and/or distribution of refugia and recolonization pathways). We should, however, keep in mind that similarities or differences in life-history traits (e.g., population sizes, generation times) of the interacting species might blur these simple predictions (Alvarez *et al.* 2010).

In the present study, we test for the first time these hypotheses in a plant-insect antagonistic and obligate interaction. Using mitochondrial markers in insects and genome fingerprinting in plants, we infer the respective phylogeographies of *A. maculatum* and its two main Psychodid visitors and further evaluate their congruence.

Material and Methods

Sampling

Plants and insects were sampled during springs 2006-2008, covering the whole range of the plant (Boyce 1993; Espíndola *et al.* 2010). Up to 10 leaves per population were sampled in 72 populations and dried in silica gel right after collection. All flower visitors were preserved in separated Falcon tubes filled with 70% ethanol. Insects were identified to the species following Espíndola and collaborators (*in press*) for 46 of the plant's sampling sites.

A maximum of three individuals per population of both *Psychoda phalaenoides* and *Psycha grisescens* were selected for further genetic analyses (Sup. Mat. 1).

Plant genetic analysis

DNA was extracted using the QIAGEN DNeasy Plant extraction kit, following the manufacturer's protocol (QIAGEN, Hombrechtikon, Switzerland). Alleles were produced with the amplified fragment length polymorphism (AFLP) technique (Vos *et al.* 1995), assessed with *Eco*RI and *Mse*I digestion enzymes and further amplified with two primer pairs (E-ACG/M-CTG and E-AGT/M-CAG). For digestion, 5µl of DNA were added to get a final volume of 20 µl, containing 1X Buffer 2, 0.1mg/ml BSA and 1 unit/µl of *Mse*I and *Eco*RI (New England Biolabs); the mix was incubated at 37°C for 2h. Adaptors were ligated to the digested products at 37°C for 2h, using 40µl of mix containing 1X Buffer T4, 0.45µM adaptor E, 0.36µM adaptor M, 0.015 units/µl T4 ligase (Promega SA, Dübendorf, Switzerland) and 20µl of digestion product. Preselective PCR was done in a 20µl mix composed of 1X buffer GoTaq (Promega SA, Dübendorf, Switzerland), 2µM $MgCl_2$, 250 µM dNTPs (Promega SA, Dübendorf, Switzerland), 0.25µM primers EA and MC, 0.025 units/µl of *Taq*-polymerase (Promega SA, Dübendorf, Switzerland) and 2µl ligated DNA. Conditions for this procedure were 2min at 94°C, followed by 29 cycles of 45secs at 94°C, 45secs at 56°C and 1min at 72°C, and finishing with 10min of final extension at 72°C. The final selective step was done in a 20µl mix composed of 1X buffer GoTaq (Promega SA, Dübendorf, Switzerland), 2µM $MgCl_2$, 250 µM dNTPs (Promega SA, Dübendorf, Switzerland), 0.4µM primers EA and MC, 0.025 units/µl of *Taq*-polymerase (Promega SA, Dübendorf, Switzerland) and 3µl of a solution 1/20 of the preselective PCR product. Conditions were 2min at 94°C, followed by 13 cycles of 30secs at 94°C, 30secs at 65°C with temperature lowering 0.7°C at each cycle and 1min at 72°C, followed by 23 cycles of 30secs at 94°C, 30secs at 65°C and 1min at 72°C, and finishing with 5min of final extension at 72°C. AFLP profiles were genotyped by Macrogen Inc. (South Korea). Allele binning was first automatically processed using Genemapper 3.7 (Applied Biosystems) by applying a RFU threshold of 50 RFU in a range of 50bp to 350bp. Bins were afterwards visually checked and adjusted when necessary.

Because digestion and amplification problems can arise during the procedure and can highly bias results (Arrigo *et al.* 2009), we randomly chose 36 individuals per plate to be used as intraplate replicates, while five samples were selected out of the total to be used as interplate replicates. The percentage of bands shared by all the replicates of the same sample was analyzed and repeatability was assessed.

In order to identify the genetic structure of *A. maculatum*, 20 runs of 1.000.000 generations with a 200.000 burnin period for each *K* prior number of populations (set between 1 and 20) were performed using a model-based assignment algorithm as implemented in Structure 2.2 (Falush et al. 2007). Following recommendations of the authors, data was coded as diploid and recessive alleles as present. The most probable *K* was identified using approaches proposed both by Pritchard *et al.* (2000) and Evanno *et al.* (2005). A non model-based algorithm was also applied to infer clustering of specimens, by using the K-means method, following Hartigan and Wong (1979), and selecting for the best number of groups using the inertia criterion (*i.e.*, the average of the distances between the centroid of each cluster and each sample contained in it), following Kergoat & Alvarez (2008). This method, recently applied in a phylogeographic framework by Burnier *et al.* (2009) and Arrigo *et al*. (2010) is based on the allele composition of each sample and allows the identification of the number of genetic clusters that optimises the grouping of samples. The analysis was performed on R 2.9.1 (R Core Development Group, 2009), applying 10.000 replicates for each *K* cluster, for values going from 1 to 20. Scripts are available upon request to the first author.

Insect genetic analysis

DNA from 152 flies belonging to the two visitor species *P. phalaenoides* and *P. grisescens* was extracted using the QIAGEN DNeasy Blood and Tissue extraction kit (QIAGEN, Hombrechtikon, Switzerland). In order to perform phylogeographic analyses, three (Cytb, COI, 16s) and one (Cytb) mitochondrial regions were amplified for *P. phalaenoides* and *P. grisescens*, respectively. PCR were done in 20µl of a mix composed of 0.5X buffer, between 1.25 and 2.5 mM $MgCl_2$, 10mM dNTP, 1 unit of GoTaq DNA polymerase (Promega,

Dübendorf, Switzerland), 0.5 µM primers (Table 1) and 3 µl DNA and run in a TGradient thermocycler (Biometra, Goettingen, Germany). The program consisted of 2:30 min at 95°C, followed by 35 cycles of 30 sec at 95°C, 40 sec at 57°C or 48.5°C (Table 1), 1 min at 72°C, and finishing with a final elongation of 8 min at 72°C. Amplified fragments were sequenced by Macrogen Inc. (South Korea) and Fasteris SA (Switzerland). Sequences were visually corrected on Chromas Pro 1.41 (Technelysium Pty. Ltd.) and further aligned on BioEdit 7.0.4.1 (Hall 1999). Gaps were coded using the method of Simmons and Ochoterena (2000) as implemented in FastGap 1.2 (Borchsenius 2009).

Phylogenetic relationships were inferred using Bayesian and Maximum Likelihood (ML) approaches. Models of evolution were estimated using MrAIC v. 1.4.4 (Nylander, 2004) and Bayesian analyses were performed on MrBayes 3.1.2 (Huelsenbeck & Ronquist 2001; Ronquist & Huelsenbeck 2003) running 50.000.000 generations, with temperature 0.5 and sampling every 1.000 generations. ML analyses were run on RaxML 7.2.6 (Stamatakis 2006) based on 10.000 bootstraps. Complementary phylogenetic networks were constructed using TCS v1.21 (Clement *et al.* 2000) applying a 95% connection limit.

In order to test for isolation by distance, correlations between genetic and geographic distances were calculated for both species. For this, considering the models of evolution previously inferred for each partition and each species, we calculated the pairwise genetic distance between samples using the *Dset* function as implemented in PAUP* 4.0 (Swofford 2003). To account for multiple partitions in *P. phalaenoides*, distances were calculated for each partition and were afterwards averaged. Geographic distances were calculated between all samples and correlations were adjusted between both matrices using GenAlEx 6.3 (Peakall & Smouse 2006).

Comparative analysis

Phylogeographic patterns for both the plant and the insects were displayed on geographical maps using ArcMap 9.3 (ESRI, Redlands CA, USA), with populations being represented as pie charts showing the number of individuals assigned to each cluster or haplotype.

Congruence was visually evaluated and was accepted if the spatial distribution and the number of genetic clusters were similar for both the plants and the insects.

Results

Phylogeography of *A. maculatum*

A total of 394 individuals were extracted, out of which 362 successfully amplified bands . Of these, 62 did not present any band with primer pair E-ACG/M-CTG, while 4 did not amplify with primer pair E-AGT/M-CAG. Finally, primer pair E-ACG/M-CTG provided 166 markers, while E-AGT/M-CAG provided 160. Reproducibilities were 0.963 and 0.957 for E-ACG/M-CTG and E-AGT/M-CAG, respectively, calculated according to Bonin *et al.* (2004).

Structure runs showed that the most likely number of genetic clusters *K* for the *A. maculatum* dataset was two. Genetic groups were geographically structured (Fig. 1a) with one group spreading through the Balkans and Italy and the other covering the rest of Europe.

The non-hierarchical K-means approach indicated that K-values providing good inertia values were three and six (Sup. Mat. 2). In the first case (*K*=3; Fig. 1b), one of the three clusters appeared geographically structured: while two admixed groups were distributed throughout the central and northern European region, the third one was restricted to the Balkans and Italy. In the latter case (*K*=6; Fig. 1c) a similar pattern of admixture among four clusters was retrieved for the central and northern European area, while in addition to the group covering the Balkans and Italy, one supplementary cluster was exclusive of the Carpathian region (pink in Fig. 1c).

Phylogeography of *P. phalaenoides* and *P. grisescens*

A total of 104 samples of *P. phalaenoides* successfully amplified three regions (1428 bp), while 45 *P. grisescens* amplified one region (679 bp) (Table 2; sequences will be submitted to GenBank during the review process).

Both phylogenetic topologies (Sup. Mat. 3) and haplotype networks (Fig. 2) indicated that no geographic structure of the genetic variation was retrieved in either of the two species

analyzed. For convenience of representation, we illustrate this result by focusing on the haplotype network approach.

In *P. phalaenoides* (Fig. 2a), 24 different haplotypes were recovered, 21 of them being unique to single specimens (white in Fig. 2a). The three others (pink, blue and green, in Fig. 2a) were present in 80, three and two specimens, respectively. While the most frequent haplotype (pink) was separated of the others by at least four mutational steps, the remaining types were more interrelated. Even if unique haplotypes appeared to be relatively more abundant in the Eastern edge of the sampled zone, two main haplotypes (pink and blue) were present homogeneously across the sampling range and the remaining haplotype (green) was present only at the Eastern edge of the distribution, but included only two samples.

In *P. grisescens* (Fig. 2b), 22 different haplotypes were retrieved, 19 being unique to single specimens (in white on the Fig. 2b). The three others (pink, blue and green) were present in 13, nine and three specimens, respectively. All 22 haplotypes were highly interrelated with 20 of them showing distances to the closest haplotype equal to one step, and only two demonstrating distances equal to two steps. Here again, the spatial distribution of the main haplotypes was not structured.

When testing isolation by distance (Fig. 3), results indicated that in both species there was no correlation between genetic and geographic distances (R^2=0.003 and 0.006 for *P. phalaenoides* and *P. grisescens*, respectively, with *P* values > 0.05 in both cases).

Comparison of phylogeographic patterns

The phylogeographic patterns of *Arum* and each of the two Psychodids presented no congruence: neither the number nor the distribution of genetic groups were similar between the interacting species. Whereas the plant showed a trend towards a spatial structure of its genetic variation, insects' haplotypes were distributed irrespective of any geographic organization.

Discussion

Phylogeography of *A. maculatum*

Our results demonstrate that *A. maculatum* shows a trend towards a spatial structure of the genetic variation (Fig. 1), with one [*Structure* and K-means (K=3) approaches] or two [K-means (K=6) approach] clusters exclusive of the Balkan and Italian region and of the Carpathians, and the others widespread throughout Europe (Fig. 1a-c). This pattern is similar to those unravelled in other European organisms, as for instance, *Ursus arctos* or *Sorex araneus* (Hewitt 1999), where one widespread European genotype was segregated from two or one clusters endemic of south-eastern Europe. The observed split of the south-eastern region into two groups (i.e., Italy and Balkans on one hand, and Carpathians on the other hand) at higher values of K (K=6) is also congruent with climatologic records, which indicate that during the Last Glacial Maximum (LGM; around 18Kya) the Northern Adriatic was dry land because of lowered sea levels and a communication between its Eastern and Westerns coasts was possible (Mussi 1990), whereas gene flow with the Carpathian area was probably more restricted.

It is worth noting that a large part of genetic variation identified by the non-model based approach seems to be little or not spatially structured. This is the case of two and four clusters in the *K*=3 and *K*=6 analyses, respectively. However, because the model-based approach (i.e., Structure) did not identify these groups, it is difficult to conclude whether the admixed pattern at the central and northern part of the European distribution area is artefactual or not. Splitting overestimation by the K-means approach could result from a main principle of the method, which aims at minimizing among-sample distances. As a consequence, the more clusters it creates, the smaller the distances between the centroid of each cluster and each included sample are, a bias similar to overfitted correlations. Following this hypothesis, the multiple central-European clusters identified would be more likely an indication of the difficulty to define a best *K* than a biological reality.

Nonetheless, if we assume that this admixture result reflects a biological characteristic, we could imagine that specimens from this region are carrying a large amount of genetic diversity present in pre-LGM populations of *A. maculatum*. In this scenario, plants from the

different genetic clusters identified in the K-means approach would have, at some time, shared Pleistocene refugia, and among-lineage cross-reproduction might have occurred. The pattern currently observed would thus be a consequence of ancient polymorphism combined with massive gene flow among lineages. After the last glacial retraction (around 13 Kya), immigrants from these refugia would have reinvaded the region, carrying in their genomes both the signature of ancient polymorphisms, but also the effect of more recent refugial genetic admixture. A result that supports this view is the pattern observed in the distribution of Nei diversities (Sup. Mat. 4), which indicates centres of high diversity in both Italian and south-eastern areas and in regions where several of these central-European clusters are shown to be admixed.

Insect Phylogeography

Despite that the markers used revealed a high level of genetic variability, with more than 20 haplotypes identified in both cases (see Table 2 and Fig. 2), no geographic structure could be retrieved in either of the species. Considering the large among-population distances (up to 3.800 km), this pattern could indicate that gene flow between far distant populations is possible in Psychodids despite the flies' small sizes, which is in agreement with the significant lack of isolation by distance (Fig. 3). Moreover, because of their high population sizes and number of generations per year, it is possible that substitution rates are extremely high (see below). Although it is empirically difficult to quantify dispersal by a direct approach in these species, *P. phalaenoides* was recently sampled in aerial netting surveys in England, where insects were collected at around 200m above ground, a height at which any insect found is considered to be engaged in migratory –and thus potentially large-scale– movement (Chapman *et al.* 2004). Moreover, in at least one previous study (Hardy & Cheng 1986), Psychodids were shown to be able to perform extended flights of at least 160km, since members of this family were collected alive from boats navigating at this distance offshore in the North Sea. Such biological evidences are highly compatible with the absence of spatial genetic structure observed in our results, in which we indirectly reached the same conclusion. It is also interesting to point out that high level of gene flow –with its associated

lack of spatial genetic structure- has been also observed in other insect species, as demonstrated in the case of *Ips typographus* (Coleoptera: Scolytidae) at a regional scale in Switzerland (Gugerli *et al.* 2008).

Another explanation for this lack of phylogeographic structure in combination with high dispersal rates could also lay in the biology of these species, which could have been associated to human activities since the Holocene. It is known that these insects are tightly related to agricultural and anthropized environments, since they develop in rich substrates as cow dung, sewage works and decaying matter (Satchell 1947; Vaillant 1971). Because of this characteristic, it is possible that a part of the mixture of genetic types could be due to both old and recent human activities, as rapid (fuelled) long distance movements currently used for the transport of cattle and waste. Such an effect of human-mediated dispersal on the spatial genetic pattern of insect species has already been observed in other taxa associated with human activities, such as the bean pest beetle *Acanthoscelides obvelatus* (Alvarez *et al.* 2007).

Comparative Phylogeography

The respective presence and absence of genetic structure in plants and insects indicate that there is no congruence between both genetic structures. Not finding any evidence of shared spatial genetic structure would confirm our first hypothesis: both species appear to present independent refugial survival and recolonization pathways. The antagonistic interaction seems thus to have persisted through Pleistocene climatic oscillations even if lineages of the independent species (*i.e.,* the insect) do not systematically find shelter in the same refugia as the dependent species (*i.e.,* the plant). However, the maintenance of the interaction, and of the dependent plant species if it harbours a low level of plasticity, requires that at least one refugium is shared by the two species. Similarly, the interaction can persist despite differences in recolonization routes, as far as the independent species is frequent enough to allow the dependent species to find both suitable abiotic and biotic habitats.

It should be noted that a recent post-glacial recolonization simulation-based study performed by Borer et al. (submitted) proposed that a pattern similar to the one obtained in

348 the present study (*i.e.,* presence and absence of genetic structure in one and the other
349 partner, respectively) could be retrieved when the two interacting species disperse at very
350 different rates. This could be the case in the interaction studied here as well, since it is
351 known that Psychodid flies can reach enormous population sizes [several thousand
352 individuals per square meter (Arshad & Moh Leng 1991)] with extremely short generation
353 times when compared to those of *A. maculatum* [12-18 generations per year in the flies
354 (Vaillant 1971) *vs.* one generation every 10-15 years in the plant (Boyce 1993)]. This can
355 allow for higher dispersal rates through time, increase the mutation rate and enhance the
356 probability of genotype admixture.

357 Despite this incongruence in spatial genetic structures, recent large-scale studies on the
358 biogeography of pollinators associated with *A. maculatum* (Espíndola *et al.* in press)
359 indicated that respective densities of the two Psychodid flies were geographically structured
360 (Fig. 1d), roughly matching the plant spatial genetic structure identified in the current study
361 (Fig. 1a-c). The relative abundance of *P. grisescens* as a visitor of *A. maculatum* is much
362 higher in south-eastern Europe, an area in which the plant is characterized by a marked
363 spatial genetic structure. If this correlation is not random, this would indicate that the spatial
364 genetic structure of the plant could be currently shaped by the spatial variation of the
365 biological characteristics of pollinators in the different regions, the main principle of the theory
366 of geographic mosaic of coevolution (Thompson 2005). Another option could be that the
367 current genetic structure of the plant reflects selective and adaptive processes related to
368 pollinator attractions having happened during the last glacial period. In the future, coalescent-
369 based hypothesis testing on AFLP data could provide a statistical answer to the question of
370 whether or not the phylogeographic structure found in the plant is defined by history (i.e., in
371 relation to its hindcasted LGM distribution) or pollination ecology (i.e., defined by the
372 Psychodid composition).

373

374 **Acknowledgements**

The authors would like to thank Dr. Y. Triponez, Dr. Z. Kącki, Dr. Z. Dajdok, Dr. J. Ježek, Dr. M. Gibernau, Dr. A. Diaz, Dr. F. Conti, P. Zryd, R. Lavigne and N. Revel for their invaluable assistance during field work and data analysis. They also thank the publishers of OIKOS for permission to use the illustration reproduced in Figure 1d. This project was funded by the Swiss National Science Foundation (project No. 3100A0-116778).

References

Alvarez N, Hossaert-McKey M, Restoux G, Delgado-Salinas A, Benrey B (2007) Anthropogenic effects on population genetics of phytophagous insects associated with domesticated plants. *Evolution* **61**, 2986-2996.

Alvarez N, Kjellberg F, McKey D, Hossaert-McKey M (2010) Phylogeography and historical biogeography of obligate specific mutualisms. In: *The Geography of Host-Parasite Interactions* (eds. Morand S, Krasnov B), pp. 31-39. Oxford University Press, UK.

Arrigo N, Felber F, Parisod C*, et al.* (2010) Origin and expansion of the allotetraploid *Aegilops geniculata*, a wild relative of wheat. *New Phytologist* **187**, 1170-1180.

Arrigo N, Tuszynski JW, Ehrich D, Gerdes T, Alvarez N (2009) Evaluating the impact of scoring parameters on the structure of intra-specific genetic variation using RawGeno, an R package for automating AFLP scoring. *BMC Bioinformatics* **10**, 33.

Arshad A, Moh Leng K-Y (1991) Preliminary population assessment of *Psychoda alternata* (Diptera: Psychodidae) in soil irrigated with Wastewater for turf cultivation. *The Florida Entomologist* **74**, 591-596.

Avise JC (2009) Phylogeography: retrospect and prospect. *Journal of Biogeography* **36**, 3-15.

Bonin A, Bellemain E, Bronken Edeisen P*, et al.* (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* **13**, 3261-3273.

Borchsenius F (2009) FastGap 1.2, Department of Biological Sciences, University of Aarhus, Denmark. Published online at http://192.38.46.42/aubot/fb/FastGap_home.htm.

402 Borer M, Buerki S, Arrigo N, Naisbit RE, Alvarez N (submitted) Comparative phylogeography
403 in a specific plant-insect antagonism: partial congruence in large-scale spatial genetic
404 structures associated with similar dispersal abilities. *Molecular Ecology*.
405 Boyce PC (1993) *The genus Arum* The Royal Botanical Gardens, Kew, London, UK.
406 Burnier J, Buerki S, Arrigo N, Kupfer P, Alvarez N (2009) Genetic structure and evolution of
407 Alpine polyploid complexes: *Ranunculus kuepferi* (Ranunculaceae) as a case study.
408 *Molecular Ecology* **18**, 3730-3744.
409 Chapman JW, Reynolds DR, Smith AD, Smith ET, Woiwood IP (2004) An aerial netting study
410 of insects migrating at high altitudes over England. *Bulletin of Entomological*
411 *Research* **94**, 123-136.
412 Clement M, Posada D, Crandall K (2000) TCS: a computer program to estimate gene
413 genealogies. *Molecular Ecology* **9**, 1657-1660.
414 D'Andrea L, Broennimann O, Kozlowski G*, et al.* (2009) Climate change, anthropogenic
415 disturbance and the northward range expansion of *Lactuca serriola* (Asteraceae).
416 *Journal of Biogeography* **36**, 1573–1587.
417 Diaz A, Kite GC (2002) A comparison of the pollination ecology of *Arum maculatum* and *A.*
418 *italicum* in England. *Watsonia* **24**, 171-181.
419 Ehlers J, Gibbard PL (2007) The extent and chronology of Cenozoic Global Glaciation.
420 *Quaternary International* **164-165**, 6-20.
421 Espíndola A, Buerki S, Küpfer P, Bedalov M, Alvarez N (2010) New insights into the
422 phylogeny and biogeography of *Arum* L. (Araceae): unravelling its evolutionary
423 history. *Botanical Journal of the Linnean Society* **163**, 14-32.
424 Espíndola A, Pellissier L, Alvarez N (*in press*) Variation in the proportion of flower visitors of
425 *Arum maculatum* across its distributional range, in relation with community-based
426 climatic niche analysis. *OIKOS*. doi: 10.1111/j.1600-0706.2010.18937.x.
427 Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using
428 the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.

429 Farrell BD (1998) "Inordinate Fondness" explained: why are there so many beetles? *Science*
430 **281**, 555-559.
431 Farrell BD, Mitter C, Futuyma DJ (1992) Diversification at the insect-plant interface. Insights
432 from phylogenetics. *BioScience* **42**, 34-42.
433 Futuyma DJ, Agrawal AA (2009) Macroevolution and the biological diversity of plants and
434 herbivores. *Proceedings of the National Academy of Sciences of the United States of*
435 *America* **106**, 18054-18061.
436 Gugerli F, Gall R, Meier F, Wermelinger B (2008) Pronounced fluctuations of spruce bark
437 beetle (Scolytinae: *Ips typographus*) populations did not result in bottleneck-induced
438 genetic differentiation. Forest Ecology and Management 256: 405–409.
439 Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis
440 program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95-98.
441 Hardy AC, Cheng L (1986) Studies in the distribution of insects by aerial currents. III. Insects
442 drift over the sea. *Ecological Entomology*, 283-290.
443 Hartigan JA, Wong MA (1979) A K-means clustering algorithm. *Applied Statistics* **28**, 100-
444 108.
445 Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and
446 speciation. *Biological Journal of the Linnean Society* **58**, 247-276.
447 Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the*
448 *Linnean Society* **68**, 87-112.
449 Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny.
450 *Bioinformatics* **17**, 754-755.
451 Lack AJ, Diaz A (1991) The pollination of *Arum maculatum* L. - a historical review and new
452 observations. *Watsonia* **18**, 333-342.
453 Mussi M (1990) Continuity and change in Italy at the Last Glacial Maximum. In: *The world at*
454 *18000 BP* (eds. Soffer O, Gamble C), pp. 126-147. Unwin Hyman, London.

455 Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction
456 endonucleases. *Proceedings of the National Academy of Sciences of the United*
457 *States of America* **76**, 5269-5273.
458 Nogués-Bravo D, Rodriguez J, Hortal J, Batra P, Araújo MB (2008) Climate change, humans,
459 and the extinction of the woolly mammoth. *PLoS Biology* **6**, 79.
460 Nylander JAA (2004) MrModeltest v2. Program distributed by the author. Evolutionary
461 Biology Centre, Uppsala University.
462 Ollerton J, Diaz A (1999) Evidence for stabilising selection acting on flowering time in *Arum*
463 *maculatum* (Araceae): the influence of phylogeny on adaptation. *Oecologia* **119**, 340-
464 348.
465 Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic
466 software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.
467 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using
468 multilocus genotype data. *Genetics* **155**, 945-959.
469 Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under
470 mixed models. *Bioinformatics* **19**, 1572-1574.
471 Satchell GH (1947) The ecology of the British species of *Psychoda* (Diptera, Psychodidae).
472 *Annals of Applied Biology* **34**, 611-621.
473 Schmitt T (2007) Molecular biogeography of Europe: Pleistocene cycles and postglacial
474 trends. *Frontiers in Zoology* **4**, 11.
475 Simmons MP, Ochoterena H (2000) Gaps as characters in sequence-based phylogenetic
476 analyses. *Systematic Biology* **49**, 369-381.
477 Soberón J, Peterson TA (2005) Interpretation of models of fundamental ecological niches
478 and species' distributional areas. *Biodiversity Informatics* **2**, 1-10.
479 Stamatakis A (2006) RAxML-VI-HPC: Maximum Likelihood-based phylogenetic analyses
480 with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690.
481 Stewart JR, Lister AM (2001) Cryptic northern refugia and the origins of the modern biota.
482 *Trends in Ecology and Evolution* **16**, 608-613.

Stewart JR, Lister AM, Barnes I, Dalén L (2010) Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society of London B* **277**, 661-671.

Swofford DL (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7**, 453-464.

Thompson JN (2005) Coevolution: the geographic mosaic of coevolutionary arms races. *Current Biology* **15**, 992-994.

Tsai Y-HE, Manos PS (in press) Host density drives the postglacial migration of the tree parasite, *Epifagus virginiana*. *Proceedings of the National Academy of Sciences of the United States of America*. **107**, 17035-17040.

Vaillant F (1971) 9d. Psychodidae-Psychodinae. In: *Die Fliegen der Palaearktischen Region* (ed. Lindner E), p. 358. E. Schweizerbart'sche Verlagsbuchhandlung, Stuttgart.

Vos P, Hogers R, Bleeker M*, et al.* (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**, 4407-4414.

Author information box

Anahí Espíndola is a PhD student at the University of Neuchâtel (Switzerland), interested in ecological, spatial and phylogeographic aspects of the evolution, establishment and dynamics of plant-insect interactions. Nadir Alvarez is a research associate at the University of Lausanne (Switzerland), interested in the evolution of plant-insect interactions at different scales of time and space.

507 Table 1 – Primer sequences and respective annealing temperatures used to sequence
508 portions of the insects' mtDNA.
509

Region	Primers	Sequence (5' -> 3')	Annealing (°C)
COI	C1-J-1718	GGG GGG TTT GGA AAT TGA TTA GTG CC	48.5
	Tl-2-N3014	TCC ATT GCA CTA ATC TGC CAT ATT A	48.5
Cytb	CB-J-11338	CAC ATT CAA CCA GAA TGA TAT TT	57
	N1-N-12051	GAT TTT GCT GAA GGT GAA TCA GA	57
16s	LR-N-13398	CGC CTG TTT AAC AAA AAC AT	57
	LR-J-12883	CCG GTT TGA ACT CAG ATC ATG T	57

510

511 Table 2 – Total, variable (V), parsimony-informative (PI) and constant (C) sites per
512 species and sequenced region, in base pairs, for the three mtDNA regions sequenced in
513 insects.

Species	Region	C	V	PI	**Total**
P. phalaenoides	16s	209	14	0	223
	Cytb	646	34	14	680
	COI	509	16	5	525
	Total	1364	64	19	**1428**
P. grisescens	Cytb	611	68	13	**679**

514

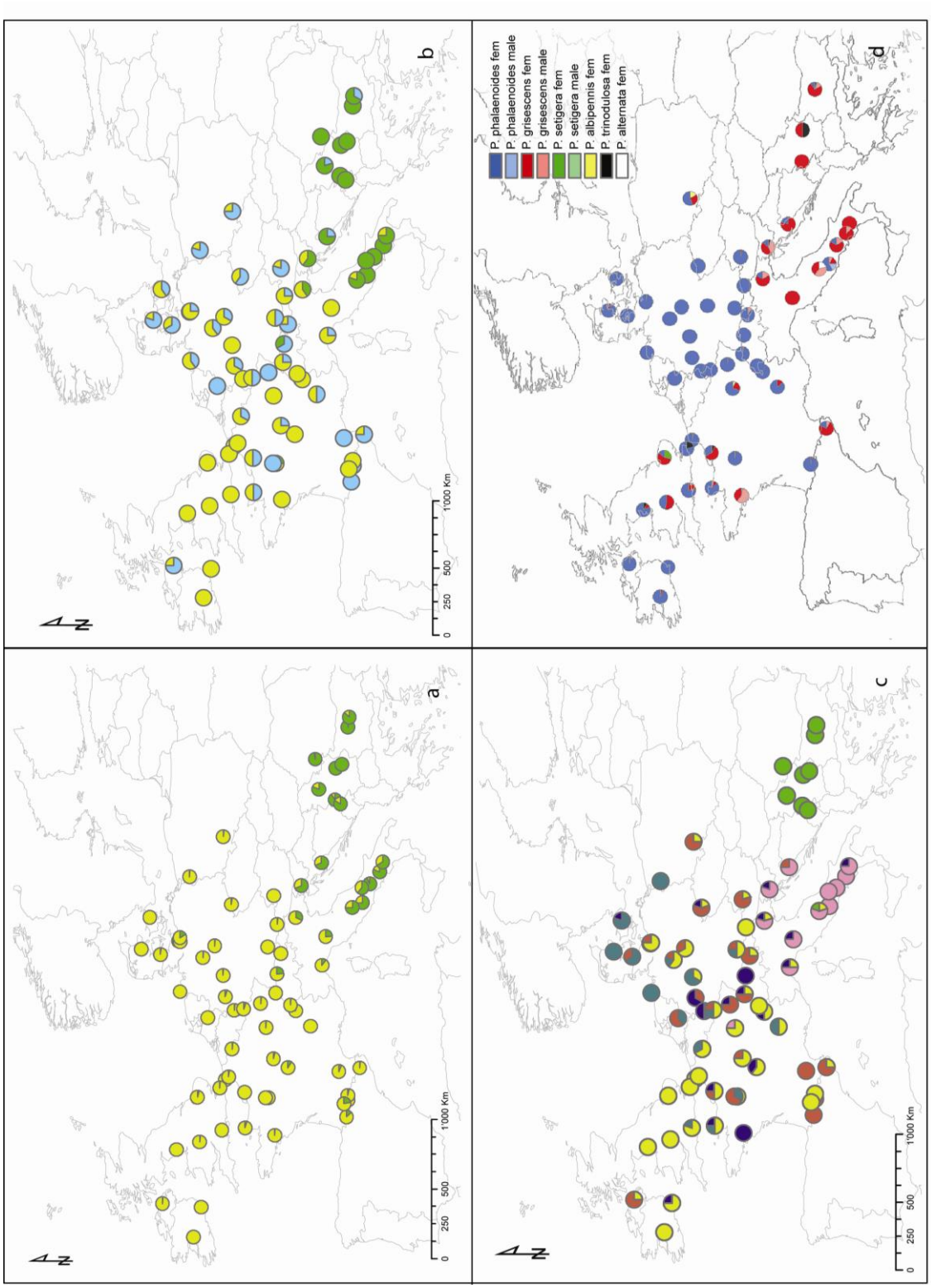


Figure 1 – a-c) Genetic clusters inferred under Bayesian (a) and non model-based (b and c) frameworks. Colours indicate clusters; portions in pies indicate probabilities of assignment (for Bayesian) or number of individuals (for K-means) assigned to each cluster. d) Composition of Psychodid visitors –and putative pollinators– of *A. maculatum* (from Espíndola *et al. in press*; reproduced with permission of the editors). Refer to figure legend for identity of species.

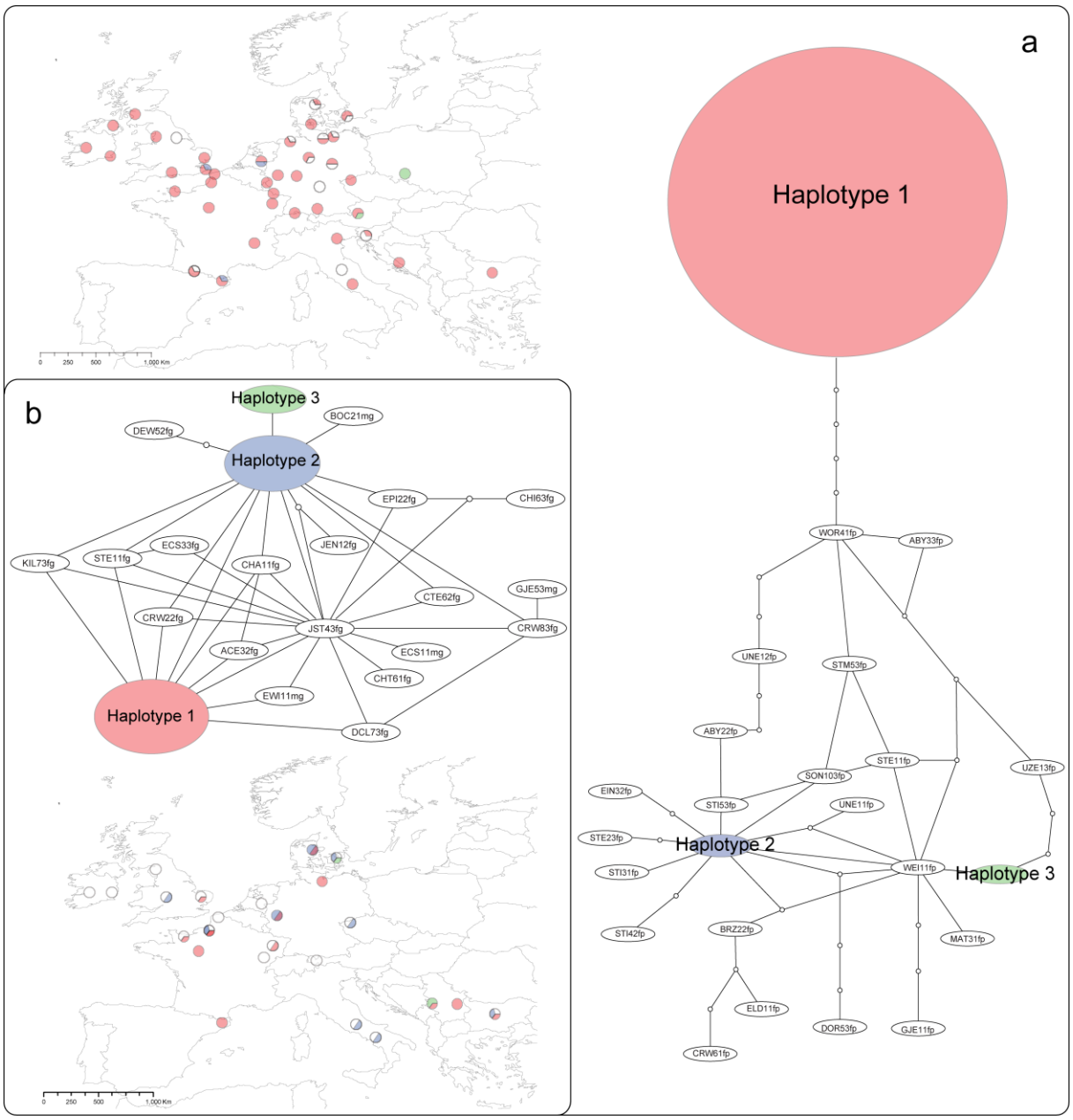


Figure 2 – Haplotype networks and respective geographic distribution of most-frequent haplotypes obtained for *Psychoda phalaenoides* (a) and *Psycha grisescens* (b).

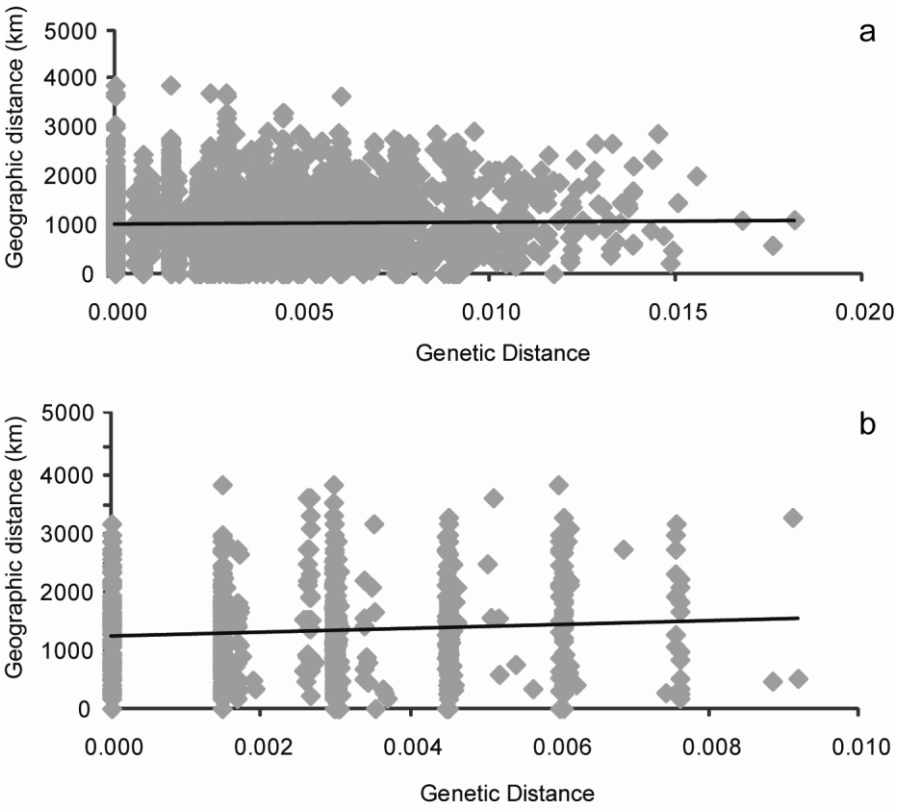


Figure 3 – Correlations between genetic and geographic distances for *P. phalaenoides* (a) and *P. grisescens* (b). Adjusted regression lines are shown (R^2=0.003 and 0.006, *P*-values: 0.350 and 0.170, respectively).

Supplementary Material 1 – List of sampled locations with geographic coordinates.

Number of plants, *Psychoda phalaenoides* and *Psycha grisescens* collected at each location.

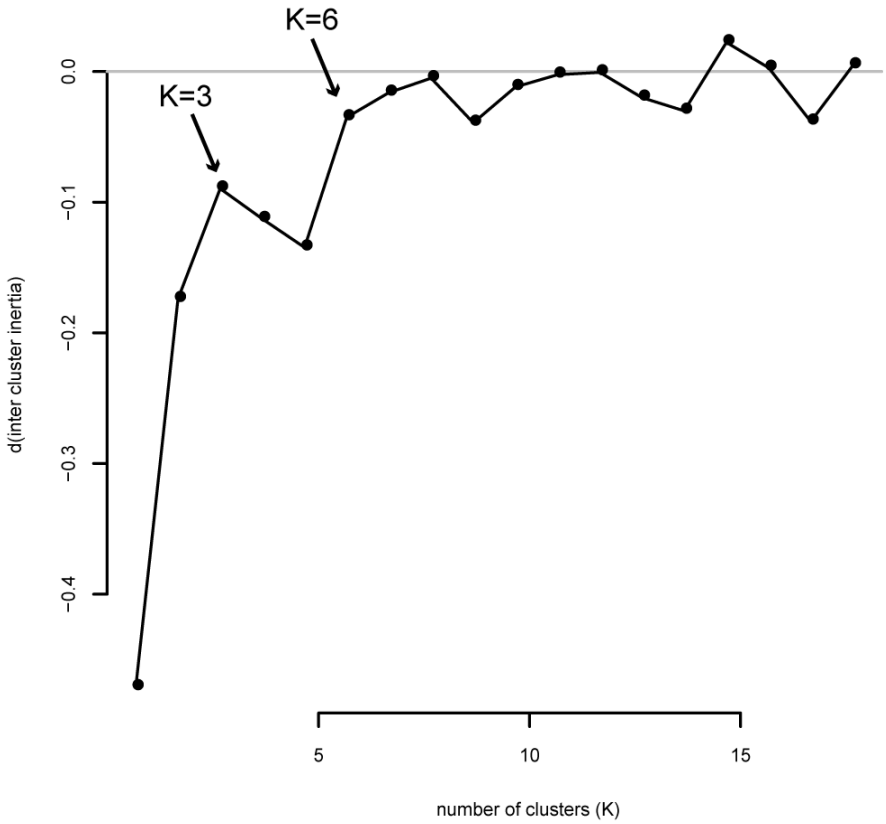
Location	Code	Longitude	Latitude	Altitude (m)	*A. maculatum*	*P. grisescens*	*P. phalaenoides*
Aby	ABY	10.16801	56.14328	0	5	2	3
Acerno	ACE	15.16854	40.77018	928	5	2	
Avala	AVA	20.51481	44.69189	468	5		
Bausen	BAU	0.7239	42.82485	610	5		
Baile Herkulane	BHK	22.4586	44.93176	300	5		
Blankenheim	BLA	7.098333	50.73389	531	5	2	3
Boc	BOC	13.06557	50.34522	364	5	2	2
Bouloire	BOU	0.56135	47.97096	115	5	1	1
Brzeg	BRZ	17.4509	50.85675	336	5		2
Bois St. Pierre	BSP	3.71207	50.29306	130	5		
Chaumont	CHA	5.09475	48.11508	296	5	1	3
Chiflik	CHI	24.52836	42.813	786	5	3	1
Churchtown	CHT	-2.78686	53.88121	3	5	1	3
Cranwich	CRW	0.60818	52.52699	24	5	3	1
Conteville	CTE	1.73872	50.73731	60	5	1	3
Duclair	DCL	0.92016	49.48517	-1	5	3	2
Dewley	DEW	-2.2918	52.38652	35	5	2	2
Dornum	DOR	7.43117	53.65095	0	5		3
Ecausseville	ECS	-1.37439	49.46218	0	5	3	2
Einhaus	EIN	10.73472	53.71355	21	5	1	3
Eldgasen	ELD	9.62534	52.17533	101	5		2
Epinal	EPI	6.654722	48.45028	400	5	2	3
Ewijk	EWI	5.74856	51.8763	21	5	1	3
Falkstone	FLK	1.19223	51.09506	78	5		3
Font de la Salut	FON	2.49728	42.06183	1100	5	3	3
Foret d'orient	FOR	0.56135	48.1518	295	5		
gabrovnitza	GAB	25.17856	42.7295	731	4		
Gross Disnack	GDI	10.70769	53.72313	42	13		
Chene-Bourg	GE	6.18533	46.192782	NA	2		
Genos	GEN	0.40551	42.81207	975	5		
Gjeddesdal	GJE	12.23064	55.60755	32	5	3	3
Gostilje	GOS	19.83549	43.65561	785	5	3	
La Groutte	GRO	2.51343	46.69078	137	5		
Jenne	JEN	13.18219	41.88142	824	6	2	2
Johnstown Castle	JST	-6.50412	52.29245	-18	5	1	3
Kamena Gora	KAG	19.56567	43.2983	1201	5		
Killaloe	KIL	-8.44438	52.7946	44	5	1	2
Laccio	LAC	9.13009	44.48817	629	4		
Lescun	LEC	-0.6709	42.91148	1000	5		

Les	LES	0.7237	42.82462	610	5		
Balzers	LIE	9.886944	47.15111	494	4		3
Hohenweiler	LIN	10.315556	48.00694	486	5		3
La Loubatiere	LOU	2.2568	43.40768	720	4		
Tintesmühle	LUX	6.223889	50.15611	305	4		3
Maiche	MAI	7.339722	47.485	613	4		
Lago Matese	MAT	14.40561	41.40639	1098	10		1
Montese	MON	10.98371	44.25523	707	5		
Morro Reatino	MOR	12.85742	42.53585	759	6		
Monte Pizi	MPI	14.15127	41.92002	925	3		
La Mignonnais	MSG	-1.86466	47.55066	16	3		
Quilen	QUI	1.91746	50.52071	112	5		
Randalstown	RDT	-6.28349	54.79447	58	5		3
Rifreddo	RIF	15.82473	40.57235	1172	5		
Col de Roanza	ROA	12.20989	46.17638	816	5		1
Romsey	ROM	-1.52498	50.96424	52	5		2
St. Fargeau	SFG	3.09205	47.62233	226	5		
Sitno	SIT	14.84701	53.03406	66	5		3
Schmilau	SMI	10.7575	53.66833	2	5		
Sokobanja Vrh source	SOK	21.88755	43.60373	844	10	1	
Sonderborg	SON	9.82699	54.91554	10	7		2
Steeg	STE	13.63328	47.61349	513	5	1	3
Sankt Margrethen	STM	11.77574	47.3825	527	5		3
Suva Planina	SUV	22.13166	43.21021	1114	4		
Udbina	UDB	15.76134	44.53128	810	6		1
Unec	UNE	14.28609	45.82845	525	5		3
Ecublens	VD	6.56637	46.52655	NA	2		
Verne	VER	5.1905	45.22675	502	5		1
Hautes Pyrenées	VZE	0.17868	43.07732	580	5		3
Weilburg	WEI	8.475556	50.88361	158	3		3
Worbis	WOR	10.38198	51.42088	370	4		2
Yutz (Thionville)	YUT	6.303889	49.55139	225	4		3
Hönggerberg	ZH	8.56637	47.40812	NA	5		3

533

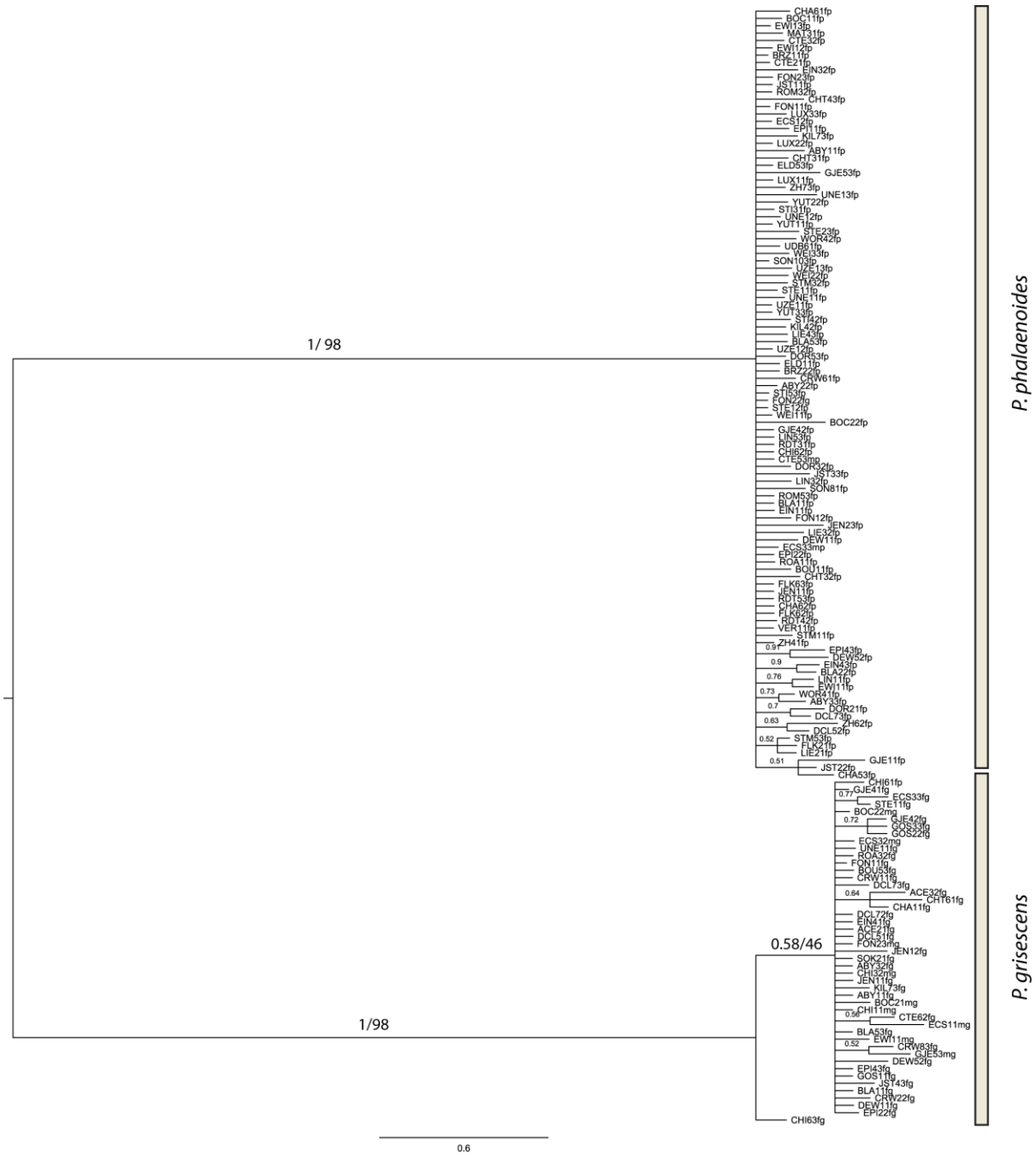
534

Supplementary Material 2. - Inertia computed, following Kergoat & Alvarez (2008), for each K value in the K-means computations on the plant AFLP dataset.



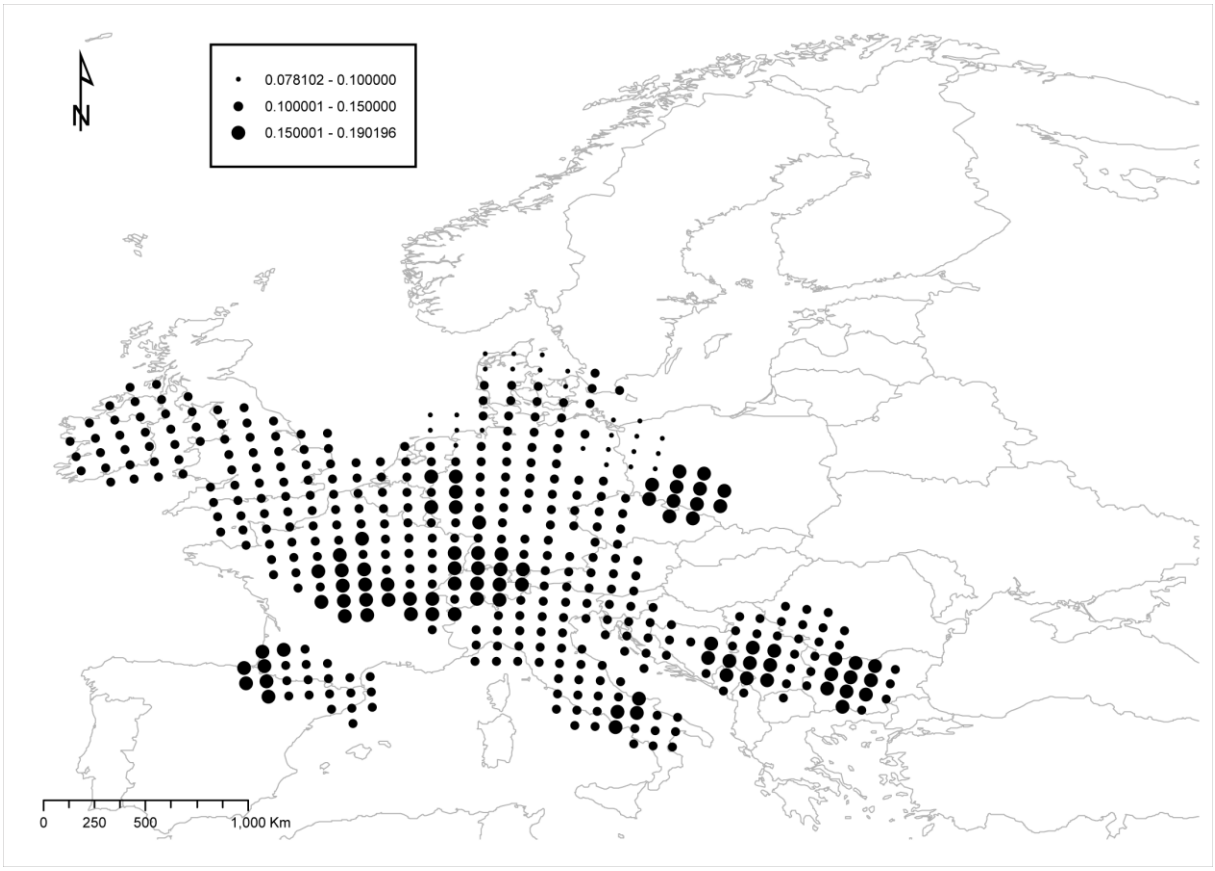
.

Supplementary Material 3 - Phylogeny of the two Psychodid sister species inferred applying Bayesian and Maximum Likelihood approaches. Support values are shown on branches, Bayesian/Maximum Likelihood.



Both species are clearly separated, but no intraspecific structure is observed in any of the species.

Supplementary Material 4 - Nei diversities of *Arum maculatum*.



Nei genetic diversities (Nei & Li 1979) in *Arum maculatum* were calculated for samples having successfully amplified fragments with both AFLP primer pairs, applying a geographic moving window run on R 9.2.1 (R Core Development Group, 2009). The method consists of the following steps: i) the region of study is defined by the most extreme sampled locations, ii) a grid of fixed cell size (in our case 75 km) and their associated centroids are calculated, iii) starting at each centroid a fixed maximum number of samples is randomly selected (three in our case) in a user-defined perimeter (150 km in this study), iv) Nei diversities are calculated for each centroid, v) points iii and iv are repeated 10’000 times, vi) mean values are calculated for each centroid. This method allows not only to calculate the Nei diversity values at each location, but also to observe their variation and distribution throughout the area of study. Moreover, it allows for a correction of eventual local sampling biases since the analysis is grid-based. Scripts are available upon request to the first author.

Chapter six

Evolutionary history of Chiastocheta species interacting with Trollius europaeus

Anahí Espíndola, Sven Buerki
and Nadir Alvarez

Evolutionary history of *Chiastocheta* species interacting with *Trollius europaeus*

Anahí Espíndola[1], Sven Buerki[2] and Nadir Alvarez[3]

[1]Laboratory of Evolutionary Entomology, Institute of Biology, University of Neuchâtel. Emile-Argand 11, 2000 Neuchâtel, Switzerland.

[2]Molecular Systematics Section, Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, United Kingdom.

[3]Department of Ecology and Evolution, University of Lausanne, Biophore Building, 1015 Lausanne, Switzerland

Abstract

The nursery pollination mutualism between *Trollius europaeus* and the species complex of pollinating flies of the genus *Chiastocheta* has been studied for several years from both behavioural and ecological perspectives. However, the species circumscription of several pollinating flies is currently not clearly established, though this is a key point for further investigating evolutionary forces driving the fate of this interaction.

Here, we aim at defining the identities and boundaries of the different *Chiastocheta* species, giving special attention to those involved in the interaction with *T. europaeus*. Based on a large-scale sampling, we apply a morphological survey and molecular-based (nuclear and cytoplasmic) phylogenetic inference to i) date the origin of the genus *Chiastocheta*, ii) infer its most likely biogeographic history, iii) define and clarify the boundaries of taxa, iv) present a model explaining congruences and incongruences between results and v) propose some guidelines for future taxonomic researches.

Our results demonstrate that the genus is relatively young (Pleistocene), and originated in Europe. Only three species (*C. rotundiventris*, *C. dentifera* and *C. lophota*) involved in the interaction with *T. europaeus* showed congruence between morphological and phylogenetic inferences. The remaining species appear to present genetic admixture, with *C. setifera* showing the highest level of hybridization.

This study demonstrates the importance of identifying potential genomic transfers between morphospecies. In the framework of the study of coevolutionary processes, our results warn researchers about the effect hybridization could have in the interpretation of ecological results, since selective forces would be probably more diffuse than previously thought, at least for those species frequently hybridizing.

Introduction

Understanding the nature of species and speciation are fundamental and much discussed topics in biology (Mallet 2010). The classical species concept is based on Linnaeus' typological philosophy (Ruse 1969), which considers species as static entities that can be described based on shared morphological characters. This sight is still of high value today, being particularly important in the field of applied conservation biology (Agapow, Bininda-Emonds et al. 2004). Other theories such as the ecological (Van Valen 1976) or biological (Mayr 1942) concepts are much more relevant from the theoretical point of view, but are also more difficult to apply for direct organismal identification. In between these two extremes is the phylogenetic species concept, which is based on molecular distances and evolutionary histories, and which under its most extreme form considers that a species should be delimited to all deeper taxa recognized in a phylogenetic classification (Mishler and Theriot 2000). Despite being sound, this concept is however also difficult to be used, because levels of divergence considered to circumscribe species are variable among biological groups (de Queiroz 2005) and because it does not necessarily allow any direct field identification of entities (a goal of diversity studies). As clearly stated by de Queiroz (2005), the fact that each concept analyzes evolutionary entities at different temporal scales, explains that taxonomic units are not necessarily congruently delimitated when different approaches are used. Moreover, most concepts generally match when dealing with well-distinguished lineages, but they are of limited application when treating complexes of species, especially if they comprise lineages with similar ecologies and/or distributions, or which are experiencing recent or ongoing genetic divergence processes. However, it is in those cases that it becomes important to adopt combined approaches to define species, since the different information each method provides could help disentangle complex situations.

Speciation processes can be influenced both by biotic and abiotic features (Ridley 1997). When factors responsible for incomplete reproductive boundaries are related with abiotic components (*e.g.*, changes in topography driving vicariance of lineages followed by quick admixture of lineages), scenarios are relatively simple to envisage. In contrast, when biotic

factors are the main drivers or limiting forces of among-lineage gene flow, understanding the evolutionary history of these lineages might require a more integrative approach. The case of the pollination mutualism between *Trollius europaeus* (Ranunculaceae) and its associated *Chiastocheta* flies species complex (Diptera: Anthomyiidae) (Pellmyr 1989) is potentially one of these biotic-influenced example. In this nursery pollination system, plants are pollinated by insects, which lay eggs on flowers and whose larvae eat the developing floral structures (Dufaÿ and Anstett 2003). Several adaptive traits are observed both in the plant and the flies: on the one hand, the plant displays flowers presenting a closed shape that excludes an important number of potential generalist pollinators and attracts high numbers of *Chiastocheta* flies, providing them a mating site and a shelter against external conditions (Pellmyr 1989). Studies have demonstrated that seed set is higher in closed than in open flowers, and that visits by non-specific visitors are less efficient than those done by *Chiastocheta* (Ibanez, Dujardin et al. 2009). It has also been observed that coloration of flowers (orange-yellow) is adapted to optimally attract *Chiastocheta* flies (Pellmyr 1989), and that the high number of carpels present in the flower appears to be related to predation costs associated with nursery pollination (Pellmyr 1992). Finally, it has been shown that *T. europaeus* produces feeding deterrent compounds that accumulate in its carpel walls, which increase in concentration when *Chiastocheta* larvae feed on the seeds (Gallet, Ibanez et al. 2007; Ibanez, Gallet et al. 2009), and which have been proposed as a regulatory method of the number of developing larvae per fruit and as a plausible starting point for explaining the high diversification (in number of species) and specialization observed in this mutualism (Ibanez and Després 2009). From the pollinators' side, fly lineages also present adaptations to the interaction, which might be a consequence of their obligate development in *T. europaeus* seeds: they display specific oviposition behaviors, with among-species differentiation on the pattern (Pellmyr 1992) and timing of oviposition (Pellmyr 1989; Pompanon, Pettex et al. 2006), and in the resource exploitation by developing larvae, probably a consequence of interspecific competition and toxin escape (Pompanon, Pettex et al. 2006).

This mutualistic system has been well studied in the European region, and it has been demonstrated that at least seven *Chiastocheta* species, described based on morphological grounds, contribute to pollination of the plant (*C. rotundiventris, C. dentifera, C. macropyga, C. trollii, C. setifera, C. inermella*, *C. lophota*) (Pellmyr 1992). Even if researchers have been working on Anthomyiidae for many years, the circumscription of fly species has been only partially recovered. First, morphological identification keys exist for eggs of all species (Pellmyr 1992), but only for some adult males and some females (Hennig 1976), while there are up to now no possibilities of morphological identification of larvae. Second, because rearing has been impossible *in vitro*, life cycles are not very well known, a situation that hinders our understanding of the biology of the species. Finally, molecular studies performed on this fly group have been focused on high taxonomic levels, comprising several –but not all– species of the genus (Després, Pettex et al. 2002). Only three species interacting with *T. europaeus* were shown to be monophyletic (*C. rotundiventris*, *C. macropyga* and *C. dentifera*), whereas the other were para-/polyphyletic (*i.e., C. inermella, C. lophota*, *C. trollii* and *C. setifera*). Moreover, some taxa (*e.g.*, *C. inermella*) showed strong evidence of gene flow or of incomplete ancestral polymorphism sorting. However, the survey lacked extensive sampling, which might have allowed more robust conclusions on the nature of species, especially considering their large distribution range (western Palearctic).

Because the system appears to present a complex history, understanding how coevolutionary dynamics shape lineage divergence in pollinating files might require a large-scale and dense sampling, together with the use of various molecular markers. This might allow recognizing and defining the associated fly species involved in this mutualism, as well as the nature and intensity of the interaction itself. Since the distribution range of the plant is important, going from the mountainous regions of southern Europe to Scandinavia, the British Islands and occupying the north-eastern Eurasian plains up to the Ural (Meusel, Jäger et al. 1965), and because the mutualism has been until now deeply studied only in the Alps and eastern Scandinavia (e.g., Pellmyr 1989; Jaeger and Després 1998; Hemborg and Després 1999; Pompanon, Pettex et al. 2006; Ibanez, Dujardin et al. 2009), we cannot

assume that the same pollinator visitation patterns will be found everywhere, as has been shown in the case of *Arum maculatum* and its fly pollinators (Espíndola, Pellissier et al. in press). The aim of the present study is thus to provide new keys for the general understanding of the identity, relationships and biogeography of *Chiastocheta* species associated with *T. europaeus*. We propose here to first analyze a reduced dataset based on samples of each morphological species (hereafter referred to as *morphospecies*) of the genus *Chiastocheta* [including also those not interacting with *T. europaeus,* as in Després *et al.* (2002)] in order to identify main lineages and clades and to infer the spatial and temporal components of the biogeographic history of the genus. Afterwards, we aim at analyzing a large-scale European sampling of fly species associated with *T. europaeus*. The combination of these two approaches will let us identify limits among taxa using molecular, morphological and spatial methods. We thus expect that the complementarity of these approaches will allow i) identifying the dynamics and timing of colonization of the main Eurasian regions by *Chiastocheta* lineages, ii) circumscribing species interacting with *T. europaeus*, iii) defining the fly species' distribution ranges, based on adult identification and iv) explaining possible reasons for barriers to hybridization between morphospecies.

Material and methods

Sampling

Fly samples covering the whole European distribution range of their associated plant *T. europaeus* were collected during springs and summers 2006, 2007 and 2008. Flies and/or larvae were sampled directly from flowers and further preserved in 70% ethanol. All insects were initially identified following Hennig (1976), which allowed a first sight on the general specific diversity in each location. Based on this raw identification, a selection of up to three individuals of the same morphospecies per sampled site was done, and determined using expert knowledge (V. Michelsen, unpublished data) even for individuals in which identification keys do not formally exist yet (*i.e.*, females in *C. macropyga*, *C. inermella, C.*

lophota; males in *C. lophota*). All double-checked samples were kept for genetic analyses and were used to produce species' occurrence maps.

Molecular analyses

Based on material obtained from the legs of insects (for adults) or from total extraction (for 25 larvae), DNA was extracted using the QIAGEN DNeasy Animal tissue extraction kit (QIAGEN, Hombrechtikon, Switzerland) following the protocol of the manufacturer. The remaining parts (abdomen, rest of thorax and head) of samples were preserved in case of further need to confirm species identification.

Nuclear (ITS) and mitochondrial (COI, COII, D-loop) regions were amplified and further sequenced. Mitochondrial data was obtained for all samples, while ITS sequencing was done only for a subset of all samples. PCR amplifications were performed using primers shown in Table 1.

All reactions were prepared in a 20 µl mix, containing 0.5X buffer, between 1 and 2.5 mM $MgCl_2$, 10mM dNTPs, 1 unit of GoTaq DNA polymerase (Promega, Dübendorf, Switzerland), 0.5 µM primers, 3 µl DNA and ran in a TGradient Thermocycler (Biometra, Goettingen, Germany). Conditions were different depending of the amplified region. For COI, COII and ITS the program started with 1:30 min at 95°C, followed by 40 cycles of 35 secs at 95°C, 1 min at 52°C, 45 secs at 72°C, and finishing with a final elongation of 8 min at 72°C. Because the amplification of the D-loop provides a longer fragment harboring A/T-rich sequences, the program consisted of 5 min at 95°C, followed by 35 cycles of 1 min at 95°C, 1 min of annealing at 55°C and 2 min of elongation at 60°C; the program finished with 5 min of final elongation at 60°C. PCR products were sequenced at Macrogen Inc. (South Korea) and Fasteris SA (Switzerland). Chromatograms were visually corrected on ChromasPro 1.41 (Technelysium Pty. Ltd.) and sequences were aligned using the Clustal-Wallis algorithm for the COI, COII and ITS regions, as implemented in BioEdit 7.0.4.1 (Hall 1999) and "moderately accurate" strategy implemented in online MAFFT version 6 (Katoh and Toh 2008; D-loop). All alignments were visually corrected using BioEdit 7.0.4.1. Gaps were coded using FastGap 1.2 (Borchsenius 2009), which uses the simple gap-coding method of

Simmons and Ochoterena (2000). Number of constant, variable and parsimony informative base pairs were calculated using MEGA 4.0 (Tamura, Dudley et al. 2007).

Divergence time estimation

Four samples from each morpholospecies associated with *T. europaeus*, and from locations distributed throughout the whole sampling range covered by this study, were chosen in order to complement those previously published by Després et al (2002). The outgroup was *Delia radicum* L. (Diptera: Anthomyiidae), following Després et al. (2002). Because we wanted to decrease the percentage of missing data as much as possible for the dating approach, we chose to work on a matrix that comprised only COI and COII regions, for which best-fit models were calculated using MrAIC (Nylander 2004) considering the Akaike's Information Criterion (AIC; Akaike 1973). Specimens used in the dating approach are listed in Supplementary Material 1.

Once the alignment was prepared, a partitioned coalescent Bayesian analysis was performed on a reduced dataset (see above for more details) with BEAST 1.5.4 (Drummond and Rambaut 2007). The COI and COII partitions were unlinked for the substitution model, whereas they were linked for the estimation of the molecular clock (see below). The best-fit model for both regions was the Hasegawa, Kishino and Yano (HKY) model with a parameter to account for a gamma among-site rate heterogeneity. One run of 100 x 10^6 generations was performed, sampling one tree every 1000 generations. Due to the lack of fossils for the genus *Chiastocheta*, direct calibration of the tree topologies was not possible and branch lengths and node ages were 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% MY-1 - 1.02% MY-1; Zakharov, Caterino et al. 2004) and 2.28% MY^{-1} (1.66% MY-1 - 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, an analysis was run

following Borer *et al.* (in press). 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 as default; Drummond, Ho et al. 2006). For each parameter, convergence of the independent 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, 95% divergence time confidence intervals were plotted on a maximum clade credibility tree using TreeAnnotator 1.5.4 (Drummond and Rambaut 2007).

Biogeographic analyses

Following recommendations of Buerki *et al.* (in press), areas were defined based on the sympatry (*i.e.*, distribution of the species) and geological criteria (sensu Sanmartín and Ronquist 2004). Four areas were thus considered (map caption in Figure 1): i) Northeastern Asia (occupying the Siberian plains and limited to the South by the mountain chains of south-eastern Russia); ii) Europe (including all the western Eurasian territory and extending as far east as the Ural); iii) Japan; iv) Southeastern Asia (the area to the south of southeastern Russian mountains). Due to computational limitations and to improve the convergence of Bayesian divergence time estimations (see above), a representative subset of the main clades was selected and biogeographic analyses were performed on the maximum clade credibility trees retrieved from the BEAST analyses. 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, Moore et al. 2005; Ree and Smith 2008) was used to investigate the biogeographic history of the *Chiastocheta* spp. 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 (Ree and Sanmartín 2009). An advantage of the DEC model is its ability to adapt a transition matrix (*i.e., Q* matrix) to reflect the changing palaeogeography, connections between areas (*e.g.*, land bridges) through time, or

dispersal capabilities of the group of interest (Buerki, Forest et al. in press). By taking advantage of this flexibility, the *Q* matrix for analyses performed in this study was constrained according to dispersal possibilities between areas (Supplementary Material 2). 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, Forest et al. in press)- it was defined to 0.01 instead of 0 when this was not true. Considering this *Q* matrix and the two dated maximum credibility clade trees obtained from the application of the different substitution rates, two biogeographic analyses were run on Lagrange v.2.0.1. Finally, following the approach described in Buerki et al*.* (in press), ancestral area reconstructions for each node were plotted on the trees using pie charts. The collection of R scripts (R Delevopment Core Team, 2009) to achieve this task are available on request to SB.

Phylogenetic inferences in Chiastocheta species associated with T. europaeus

Because topological searches were time- and resource-consuming (*i.e.*, due to the large number of samples and the length of alignments), analyses were performed at different levels and following complementary strategies. First, a dataset including all mitochondrial sequences of the 399 *Chiastocheta* samples (see above) was analyzed under both Maximum Likelihood (ML) and Maximum Parsimony (MP) approaches. It was for this dataset impossible to perform Bayesian inferences, since analyses did not reach the required parameter convergence. Second, a dataset including all nuclear sequences (237 samples) was analyzed based on the ML, MP and Bayesian algorithms. For Bayesian inference, the ITS model of evolution was estimated using MrAIC (Nylander 2004).

The ML searches were done using RAxML 7.2.6 (Stamatakis 2006) with 10’000 rapid bootstrap analyses followed by the search of the best-scoring ML tree in one single run. The three mtDNA regions were considered as one single partition (by applying the predefined GTR+G model). Analyses were done at the CIPRES cluster (San Diego CA, USA).

Non-model based analyses in MP were performed using PAUP* 4b10 (Swofford 2003), applying the parsimony Ratchet approach (as implemented in PAUPrat; Sikes and Lewis 2001) and based on ten independent searches of 200 iterations, as recommended by Nixon

(1999). For mitochondrial data, because searches identified only one most parsimonious tree, we worked on the –unique- best tree, which had the advantage of carrying information on branch lengths. For nuclear data, a majority-rule consensus topology was calculated from the most parsimonious trees. Decay indices [or Bremer supports, (Bremer 1988)] were calculated based on this topology, as implemented in TreeRot 3.0 (Sorenson and Franzosa 2007). Decay indices indicate the number of mutational steps needed to collapse a given node, and thus indicate the strength of the clade.

Bayesian analyses were performed only for nuclear data (see above) on MrBayes 3.1.2 (Ronquist and Huelsenbeck 2003), which implements Monte Carlo Markov Chain (MCMC) searches of topologies and parameters for which the data fits best. Six cold and six hot chains were run for 50'000'000 generations in two independent runs. Convergence was accepted when standard deviations reached values below 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 – checked on TRACER v1.4 (Rambaut and Drummond 2004). Topologies were thus calculated applying a burnin of 5'000'000 generations (considering thus a total of 270'006 trees) and a 50% majority-rule consensus tree was calculated.

Because one of our aims was to test the phylogenetic basis of each morphospecies, congruence between taxonomic identifications (*i.e.* based on genital characters) and monophyly was done by analyzing the position of morphospecies in the phylogeny.

In order to identify an eventual geographic structure of the different clades, their spatial distribution was plotted on maps of the region, by exploiting the geographic coordinates associated to each sample. This was done using ArcMap 9.3 (ESRI, Redlands CA, USA).

Results

Distribution of morphological species

1611 insect samples (196 larvae and 1415 adults) were collected and identified at 53 locations from the whole distribution range of the plant. We double-checked a total of 372

288 samples, among which 72 were *C. rotundiventris*, 67 *C. inermella*, 54 *C. trollii*, 52 *C. setifera*,
289 49 *C. lophota*, 41 *C. macropyga* and 38 *C. dentifera*. Along with 27 larvae, we performed
290 molecular analyses on a total of 399 samples (Supplementary Material 3).
291 Distribution ranges of morphologically identified species corresponding to the 373 adults
292 analyzed are depicted in Supplementary Material 4. While some species are distributed over
293 a wide part of the sampled area (*C. rotundiventris, C. inermella, C. setifera, C. dentifera, C.*
294 *trollii*), others present either fragmented distribution ranges (*C. macropyga*) or are restricted
295 to a specific area of the sampled zone (*i.e.*, *C. lophota* covers the southern part of Europe).
296 ***Divergence time estimations and biogeographic history***
297 47 samples provided 1134bp for the mtDNA-based analysis at the whole genus level.
298 Gaps were retrieved neither in COI nor in COII.
299 Because dating was done applying two independent substitution rates, ages of inferred
300 nodes differed between analyses (Figure 1). The origin of the *Chiastocheta* genus thus goes
301 back as far as 2.1My or 830 Ky, when applying a mutation rate of 2.28% or 0.9%,
302 respectively. Phylogenetic reconstruction inferred using both mutation rates provided
303 congruent topologies, which only differed in the position of the clade comprising *C. pellmyrii*
304 and *C. curvibasis*, two species not associated with *T. europaeus* (Pellmyr 1992).
305 Biogeographic ancestral reconstructions indicated that the most probable ancestral area
306 at the origin of the genus was Europe, with different numbers of identified dispersal events
307 as a function of the mutation rate applied. When setting a low substitution rate (0.9%) we
308 inferred five probable dispersal events (*i.e.*, three from Europe to Northeastern Asia, one
309 from central Asia to Japan and one from Northeastern Asia to eastern Asia). Alternatively,
310 we found four likely dispersal events when fixing a higher mutation rate (2.28%) (*i.e.*, two
311 from Europe to northeastern Asia, one from northeastern Asia to Japan and one from
312 northeastern Asia to eastern Asia).
313 ***Phylogenetic analyses of the dataset comprising Chiastocheta spp. associated***
314 ***with T. europaeus***

399 and 237 samples provided 2304bp for mitochondrial and 441bp for nuclear regions, respectively. While nuclear data contained seven gaps, mitochondrial sequences comprised 237 gap positions (*i.e.*, all in the D-loop region). Numbers of constant, variable, parsimony-informative and total base pairs are presented in Table 2. While all mitochondrial regions had a high level of variability (up to 44.05%), the nuclear region presented a particularly low number of variable (3.63%) or potentially parsimony informative (2.49%) sites. The best model of evolution suggested by the Akaike's Information Criterion (AIC) for ITS was F81 (Felsenstein 1981).

As shown in Figure 2, topologies obtained based on nuclear and mitochondrial data were only partially congruent (Table 3a; see below), and none showed a high level of compatibility between retrieved clades and morphospecies (Table 3b and c).

The nuclear topology (Figure 2a, Table 3b) showed the presence of at least five groups. Only one clade ("rotundiventris", ML bootstrap value = 96; Bayesian Posterior Probability = 1) fully corresponded to the morphospecies *C. rotundiventris*, including all and exclusively samples having been identified as this species. Clade "trollii" (only supported in the ML topology; Bootstrap value = 84) comprised all but two samples of those assigned to *C. trollii*, as well as four *C. setifera*. Clan "nI" included all samples of *C. dentifera*, *C. macropyga*, all but two of *C. inermella* and around one quarter of all *C. setifera* samples. Clan "lophota" contained all but one *C. lophota* and one *C. setifera*. Clade "setifera" (ML bootstrap value = 91, Bayesian Posterior Probability = 0.91, MP Decay Index = 1) was mainly formed by *C. setifera*, even if it also included two samples of *C. inermella*.

Topologies inferred from mitochondrial data (Figure 2b, Table 3c) indicated that there were five clades or groups, more or less supported depending on the analytical approach. *C. rotundiventris* was well identified as a genetically different group (ML bootstrap value = 99), completely included in clade "rotundiventris" and comprising also 17 of the larvae analyzed. The same was true for *C. dentifera*: clade "dentifera" harbored all but one sample of this species (ML bootstrap value = 100); and to a lesser extent for *C. lophota*: all but four samples of this species were included in clade "lophota" (ML bootstrap value = 100), which

also included one larva. The two remaining clades comprised a mix of the other species: on the one hand clade “mtI” (ML bootstrap value = 91) included samples belonging to different morphospecies (mainly *C. setifera*, *C. trollii* and *C. inermella*) as well as three larvae. Group “mtII” included all *C. macropyga* but also presented an important mixture of morphospecies, especially *C. inermella*, *C. setifera*, and to a lesser extent *C. lophota*.

Distribution of the different mitochondrial and nuclear clades, in relation to morphospecies are shown in Figures 3 and 4. Because in the cases of *C. rotundiventris* (Figure 3a)*, C. dentifera* (Figure 3b) and *C. lophota* (Figure 3c) mitochondrial clades and morphospecies presented similar compositions, their distributions were the same. While *C. rotundiventris* is widespread across Europe, *C. dentifera* is not present in the British Isles, and *C. lophota* is only present in southern mountain massifs.

Mitochondrial clade “mtI” (Figure 4a-c) is present in the centre and north-west of the sampled region, and is especially absent of the Balkans. Samples from this clade were distributed in three nuclear clades (Table 3a) and included the totality of clade “trollii” (Figure 4a), some individuals from “nI” (Figure 4b), and a portion of samples from clade “setifera” (Figure 4c). The latter is restricted to the (central and Eastern) Alps, and includes all *C. setifera* morphospecies from this region. The only two Scandinavian *C. setifera* samples present in this cytoplasmic clade are included in “trollii”.

Mitochondrial group “mtII” (Figure 4d-f) had representatives in almost every sampled site, and was the only homogeneously occupying the Balkans. Samples from this clade were included in geographically structured nuclear clades (Table 3a). Clade “nI” (Figure 4e) was present in Scandinavia, central and Eastern Europe, and in one population of the British Isles. Nuclear clade “setifera” (Figure 4f) is on the other hand distributed in the Southern regions and is clearly dominant in the British Isles. Only two sites presented samples included in nuclear “lophota” clade (Figure 4d). Finally, *C. macropyga* is completely included in clade “nI” (pink, Figure 4e).

Discussion

While topologies retrieved with the large dataset (*i.e.*, comprising 399 *Chiastocheta* samples from seven species; Figure 2) were compatible at the level of main clades whatever phylogenetic criterion was applied (*i.e.*, differences in the relationships among main clades are only due to one single polytomy in the ML analysis), the topology retrieved in the Bayesian inference analysis involving 47 representatives from twelve *Chiastocheta* species (Figure 1) was substantially different, probably because this analysis comprised more taxa but less mtDNA regions (only COI and COII were analyzed). However, despite the position of four out of the five main clades was different, all topologies concurred in placing *C. rotundiventris* as the most basal lineage. Relationships among the other four main lineages thus require further analyses of other DNA regions. Nonetheless, it is worth noting that several similarities were found when this topology was compared with that obtained by Després et al. (2002): *C. rotundiventris* appears monophyletic and basal, while *C. dentifera, C. latispinigera, C. curvibasis* and *C. pellmyri* are monophyletic. In contrast, we could not retrieve the monophyly of *C. macropyga*, indicating that an increase in the number of samples per species appears to have an effect on the retrieved level of taxon monophyly. *Chiastocheta lophota*, *i.e.*, the only species lacking in previous phylogenetic surveys of the genus, was placed in our chronogram as the sister lineage of *C. latispinigera* (Figure 1). Furthermore, considering the important genital morphological similarities, it is highly likely that the sample identified as C. *inermella* (inermella4) by Després et al. (2002) belongs in fact to *C. lophota.*

Divergence time estimations

Recent studies (Papadopoulou, Anastasiou et al. 2010) have pointed out the importance of the use of molecular clocks that take into account the heterogeneity of substitution rates among markers when performing temporal inferences. Because we did not have the possibility to use fossil data to calibrate our analysis, we applied two extreme substitution rates obtained from recent fossil-based molecular clock calibration studies (i.e., Zakharov, Caterino et al. 2004; Wahlberg 2006). Even if doing so did not allow us to consider variation between markers, it provided a good way to compare two extreme biogeographic and dating

scenarios for the genus *Chiastocheta*. Our results thus associate nodes to a range of ages that varies between the two analyses, but which place the root of the genus between the Ionian Middle Pleistocene (around 830 Kya) and the Gelasian Lower Pleistocene (around 2.1 Mya) for high (2.28%) or low (0.9%) mutation rates, respectively (Figure 1). Irrespectively of the kind of substitution rate used, our results concurred in placing the origin of genus *Chiastocheta* sometime during the Pleistocene, indicating that it is relatively young. A similar conclusion was also reached by Després et al. (2002) when applying a standard molecular clock (Brower 1994) which estimated the origin of the genus at 1 Mya.

Biogeographic history of genus Chiastocheta

Biogeographic analyses in a Lagrange framework (Ree, Moore et al. 2005; Ree and Smith 2008) consider branch lengths for the estimation of the most probable ancestral area. Practically this means that vicariance, extinction and dispersal events have more chances to happen when inferred on long rather than on short branches. In this study, while analyses based on the two different mutation rates provided very similar inferences of ancestral areas, they informed on different numbers of dispersal events. Our inferences (Figure 1) indicate that an early event of colonization by a lineage ancestral to *C. rotundiventris* appeared to have invaded Asia and to have diversified in that region during the Calabrian age (1.8 - 0.7 Mya). Later (between the Calabrian and Ionian ages; 1.8 – 0.12 Mya), while a European lineage survived in Europe, another (or two others in the 0.9 % analysis) dispersed and established in northeastern Asia, and from there dispersed to southeastern Asia and Japan afterwards.

The genus appears to have diversified 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 ages, and arctic species appeared to be much more widespread than today (e.g., Espíndola et al., submitted; Brochmann, Gabrielsen et al. 2003). It is possible that during these distributional expansions experienced throughout glacial periods, flies of this genus could have been able to disperse eastwards and lineages might have been isolated through vicariance processes.

Phylogeny, distribution of clades in relation to morphospecies and reinterpretation of species

A comparison of species assignments in both mtDNA and ITS analyses indicated some similarities in the phylogenetic inferences, but also several substantial differences. The main common pattern was that both analyses placed *C. rotundiventris* as a clearly defined entity (Figure 2, Table 3). It is important to note that all samples having been identified as belonging to this species were included in this supported clade. This clear molecular assignment could be explained by an early divergence in the genus (Figure 1), which allowed increasing reproductive barriers with other lineages. From a morphological point of view, however, *C. rotundiventris* should be sister to *C. abruptiventris*, its supposed Scandinavian vicariant (Pellmyr 1992). Nonetheless, even if morphological differences have been found between the two entities (V. Michelsen, unpublished data), our phylogeny provided no support to this taxon (no special clustering of samples from Scandinavia). Based on this, it seems thus more likely that the identified morphological differences represent local variations/adaptations eventually related to phenotypic plasticity.

The nuclear-mitochondrial congruence presented in the case of *C. rotundiventris* has been observed in none of the other clades/morphospecies, indicating that either an important rapid radiation in this group is happening, or/and there is/has been intensive hybridization between several lineages.

Among those species presenting incongruence between morphological and molecular identities, an important proportion of samples identified as *C. lophota* have also been grouped by molecular analyses (yellow, in Figure 2, Table 3): 92% of samples described as *C. lophota* are included in the “lophota” mitochondrial clade, while only two samples of morphospecies *C. setifera* and *C. dentifera*, respectively are also contained by this clade. Moreover, all *C. lophota* samples were included in the “lophota” nuclear clan; however, this clan also comprises some samples of other species (two *C. macropyga*). Even if the nuclear phylogeny did not provide any clade harboring exclusively *C. lophota*, it seems that either the age of the group (this clade is the third oldest in the genus; Figure 1) or/and its

455 geographic isolation (the taxon is currently only found in southern European mountains)
456 might have allowed genetic isolation despite rare events of hybridization.
457 Another well-supported mtDNA clade was the one comprising virtually only and almost all
458 (97%) *C. dentifera* morphospecies (green in Figure 2, Table 3). All samples from this clade
459 fell in the "nI" nuclear group. It is interesting noting that the mtDNA clade comprising *C.*
460 *dentifera* samples follows temporally that of *C. rotundiventris.* Similar to the treatment of *C.*
461 *lophota*, we consider that there was enough morphological (Hennig 1976), phylogenetic
462 (Figure 1 and 2) and behavioral arguments (i.e., this is the only species laying eggs in
463 central carpels at the end of flowering, Pellmyr 1989) to accept the consistency of this
464 species, despite rare occurrences of hybridization.
465 Mitochondrial "clade mtI" included samples of all species (blue-purple in Figures 2 and
466 4a-c, Table 3), except *C. rotundiventris* and *C. macropyga*. All but one *C. trollii*, 79% of *C.*
467 *inermella*, 52% of *C. setifera* and one sample of *C. dentifera* and *C. lophota* were included in
468 the group. A remarkable point is that all *C. trollii* samples of this clade (93% of all those
469 analyzed with nuclear markers) were restricted to nuclear clade "trollii" (purple in Figure 2a),
470 which indicates some genetic consistency of this morphospecies, at least from a nuclear
471 point of view, and which would provide some support to the validity of this taxon.
472 Contrastingly, an important number of the Alpine *C. setifera* included in this clade (black in
473 Figures 4d-f), were placed in nuclear clade "setifera", while the remaining were included in
474 clan "nI". It seems thus correct to consider that all *C. setifera* samples included in clade
475 "setifera" are morphospecies carrying non-hybridized genomes, while the remaining
476 probably resulted from crosses between *C. setifera* and other species (see below). It is
477 plausible that at some point in the history of diversification of the species, *C. setifera* was
478 isolated of all other taxa, which allowed a certain differentiation. It could be possible that this
479 diversifying group came afterwards into contact with other species, as *C. trollii* and *C.*
480 *inermella*; hybrids could be today represented by all samples morphologically identified as *C.*
481 *setifera* but included in clan "nI" and clade "trollii". However, because of the extremely high
482 levels of admixture, the identity of *C. setifera* should be redefined, at least from a

morphological point of view. While doing so, it would be highly recommended to consider especially those samples having been identified as this species *and* being included in the "setifera" nuclear clade; since it is likely that they are harboring non-hybridized genomes.

Another interesting point concerning clade "mtI" is that an internal subclade harboring exclusively Scandinavian *C. inermella* samples was observed in the ML approach (Figure 2b and Figure 4e). Even if this clade had a support value falling below 50%, it is worth noting it, since i) these samples also clustered together in the MP approach, and ii) this is one of the only subclades of "mtI" containing exclusively samples of one morphospecies. If this clade was correct, it would validate the formerly proposed idea of *C. inermella* as an exclusive Scandinavian species (Pellmyr 1992). It is however important to realize that nuclear results did not confirm this point (Figure 2a), which is why we recommend the redefinition of this entity.

Mitochondrial clan "mtII" (Figure 4d-f, Table 3) contained all *C. macropyga* but also samples assigned to morphospecies *C. inermella* (16%), *C. setifera* (28%) and *C. lophota* (3%) indicating that this group is far from being homogeneous. Despite this admixture, when analyzed with nuclear markers, all samples identified as *C. macropyga* were comprised in nuclear clan "nI" (pink in Figure 2a, Figure 4e, Table 3), while all remaining *C. lophota* were included in clan "lophota" (Figure 4d), and all *C. setifera* and two *C. inermella* were comprised in clade "setifera" (Figure 4f). This could indicate that in this clan *C. macropyga* is the most consistent genetic entity, clustering uniformly with both types of genetic markers (Table 3). On the other hand it appears, as mentioned above, that *C. setifera* and *C. inermella* are experiencing genetic admixtures, which would confirm our doubts about their identities.

Genetic boundaries of morphospecies

Considering the complementary phylogenetic (Figure 2), spatial (Figures 3 and 4) and morphological (Table 3b and 3c) approaches applied to the analysis of the *Chiastocheta* species complex associated to *T. europaeus*, we propose here some ideas on the genetic interspecific dynamics of the group (Figure 5). We find three different situations: i) species

511 are well defined and no hybridization is observed (*C. rotundiventris*); ii) morphology and
512 phylogenetics are congruent, but some rare hybridization events are observed (*C. dentifera*
513 and *C. lophota*); iii) hybridization is observed, which can be explained by spatial and
514 ecological reasons; species borders are in such cases less clearly circumscribed (*C.*
515 *macropyga, C. trollii, C. setifera* and *C. inermella*).

516 *Well defined species, no hybridization*

517 There are no doubts about the delimitation of *C. rotundiventris* (Figure2a-b). It has been
518 demonstrated that this species is the most basal in the genus (Figure 1 and Després, Pettex
519 et al. 2002), which indicates that the genetic distance with other lineages is the largest,
520 which could have consequences on limiting genetic compatibility with other species.
521 Moreover, behavioral studies have demonstrated that this species visits flowers significantly
522 earlier than the other species (Després and Jaeger 1999), which could probably define a
523 temporal barrier for mating. Finally, there are clear male and female genital morphological
524 differences between this and the other species (Hennig 1976, V. Michelsen pers. com.),
525 which could represent a physical barrier to hybridization.

526 *Well defined species, rare hybridization events*

527 The specific identity of *C. dentifera* is relatively clear, especially regarding the mtDNA
528 phylogeny (Figure 2b). An important amount of samples (97%) identified as *C. dentifera* falls
529 in the same exclusive mitochondrial clade. The differential genetic identity of this species
530 could be explained both by historical, morphological and behavioral reasons. Previous works
531 (Després, Pettex et al. 2002) and this study demonstrate that this species diverged early,
532 which could indicate that it had been evolving for enough time before other *Chiastocheta*
533 lineages radiated, which has led to a high level of genetic isolation. Moreover, behavioral
534 studies have also demonstrated that this species visits the flower at the very end of the
535 flowering time (Després and Jaeger 1999), when other species have generally already
536 visited *T. europaeus* inflorescences, which could directly diminish the probability of mating
537 with other species and thus reduce hybridization rate. Finally, even if there are genital
538 differences between this and the other species, they are not as important as those presented

by *C. rotundiventris*, which could also explain that some hybridization seems to happen (Figure 2 and Figure 4b). It is in our point of view very probable that the largest barrier to hybridization is in this case ecological/genetic and not necessarily morphological. However, in order to confirm this, developmental studies and hybridization tests should be performed on this species.

The case of *C. lophota* is similar to that of *C. dentifera*, with the difference that very little is known in what concerns flower visitation patterns. We can however confirm that this species is not *C. inermella*'s vicariant as previously proposed (Pellmyr 1992). Even if our study has not investigated behavioral topics, we imagine that there should be some behavioral differences that would enhance the separation of species, especially considering that morphological genital differences with non-*rotundiventris* types are not very important (Michelsen V., unpublished data), and that, despite sharing an important part of its distribution range with other species, it only rarely hybridizes.

Morphological differences, hybridization observed

This category corresponds to *C. macropyga, C. trollii, C. setifera* and *C. inermella*. Among these four taxa, those harboring stronger phylogenetic support are *C. macropyga* -completely included in groups "mtII" and "nI"-, and *C. trollii* -found only in clades "mtI" and "trollii". Interestingly, both *C. macropyga* and *C. trollii* harbor very distinctive genital differences (males of *C. macropyga* present big surstilii; females of *C. trollii* have small and short ovipositors, morphologically closer to *C. rotundiventris* than to the others; Hennig 1976).

One might ask whether or not it was possible to predict hybridization among these four morphospecies. From a morphological point of view, females among three out of the four types (*C. setifera, C. inermella* and *C. macropyga*) are similar, while males of three of them have similar types of genitalia (*C. setifera, C. inermella* and *C. trollii*) (Hennig 1976; V. Michelsen, pers. comm.). This morphological similarity could, if not explain hybridization, at least allow it, since no important physical barriers to mating would be present between morphospecies. From a genetic point of view, the four species appear closely related in the

two mitochondrial clades “mtI” and “mtII” (Figure 2), indicating that speciation processes could be recent, and that genetic incompatibility and/or differentiation is probably still an ongoing process.

A coupled analysis of nuclear-cytoplasmic incongruences (Figure 2, Table 3) and of the phylogeography of the four morphospecies (Figure 4) informs us on the direction and type of crossings that seem to be happening in the group. This combined analysis indicated that hybridization does not appear to occur at the same rate among all species (see below). Based on this, we defined a simple model (Figure 5) which takes this fact into account and explains the incongruences observed in our results.

The presence of *C. setifera* in both mitochondrial groups “mtI” and “mtII” (Figure 2b), associated with its wide distribution range (Figure 4), suggests that this species is hybridizing with the other four, and that in some cases it is also acting as a “genome carrier” between the remaining species (especially demonstrated in an internal clade of clan “mtII” in Figure 2b, where no admixture is observed between *C. macropyga* and *C. inermella*). Moreover, the only one among the four species showing some geographic structure is *C. setifera* (*i.e.,* all Alpine samples of nuclear “setifera” are in clade “mtI”, while all non-Alpine are in clade “mtII”; Figure 4) which indicates that this species is the only one observed in two mitochondrial groups. Based on these arguments we propose here two hybridizations “types”, which seem to be concerning two different *C. setifera* lineages: one Alpine (“mtI”) and one non-Alpine (“mtII”).

The Alpine lineage hybridizes with species available in that region: *C. inermella* and *C. trollii* (Figure 4a-c and Figure 5). It is interesting noting that *C. trollii* was very well identified by nuclear markers and is highly represented in the Alpine zone, while this is less the case for northern locations, which probably provides a higher probability of hybridization in the Alps than in northern areas.

On the other hand, the non-Alpine lineage of *C. setifera* is only present in “mtII”, where it is admixed with *C. inermella* (Figure 4d-f), an entity much more widely distributed here than in the Alps (Figure 4 and 5). It also admixes here with *C. macropyga*, but considering the

homogeneity in the distribution and genetic assignments of morphospecies of this species (Table 3, Figures 2 and 4), this seems to be happening at a lower rate, probably due to the more important genital differences found between species. Moreover, the absence of *C. inermella* in the two internal –non supported- clades of "mtII", could indicate that no gene flow is happening between this and *C. macropyga*, or that if some gene flow is present it happens indirectly, through *C. setifera* (Figure 5).

The species concept in Chiastocheta

As suggested by de Queiroz (2005), it appears that the example of *Chiastocheta* species represents one of those puzzling taxonomic cases in which speciation and diversification are still going processes, and that depending on the approach applied (*i.e.,* morphology, ecology, molecular differentiation), the delimitation of entities varies. In this study however, we demonstrate that the simultaneous evaluation of complementary approaches allows the definition and a better understanding of the speciation process happening in *Chiastocheta* flies associated with *T. europaeus*.

Based on our analyses we recommend that the classic systematics of the group is maintained for all species, besides *C. setifera* and eventually *C. inermella*. In spite of this, future ecological studies should not ignore the fact that genetic limits of typological species (*i.e.*, morphospecies) are not as strict as formerly supposed, which could have a non-negligible influence on the interpretation of ecological and coevolutionary studies.

Moreover, as demonstrated in this survey, there is a spatial effect in the genetic dynamics of at least one morphospecies (*C. setifera*). This effect should also be tested at an ecological level, since it is also possible that isolation and especially long-term cohabitation with different sets of species could induce taxa to behave differently in different geographic areas. In order to better understand such interspecific interactions, it is necessary to investigate behavioral features of all species and at different localities of their distribution ranges. It is possible that once this is done, we will be able to better understand the variation of the relationship at a spatial scale, not only in terms of number and identities of morphological species (as presented in this work), but also in terms of ecological functions.

Acknowledgements

The authors would like to thank the field and technical assistance of Dr. Y. Triponez, R. Lavigne, L. Juillerat, R. Arnoux, N. Villard, N. Magrou, D. Gyurova, Dr. P. Lazarevic, Dr. M. Ronikier, Dr. H. Hipperson, T. Suchan, N. Russel, P. Bowler, R. Dafydd, Dr. M. Howe and Dr. V. Michelsen. This project was funded by the Swiss National Science Foundation (project No. 3100A0-116778).

References

Agapow, P. M., O. R. Bininda-Emonds, et al. (2004). "The impact of species concept on biodiversity studies." The Quarterly Review of Biology **79**(2): 161-79.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. Second International Symposium on Information Theory. A. Kiado. Budapest**:** 267-281.

Borchsenius, F. (2009). FastGap 1.2. Department of Biological Sciences, University of Aarhus, Denmark, Published online at http://192.38.46.42/aubot/fb/FastGap_home.htm.

Borer, M., N. Alvarez, et al. (in press). "The phylogeography of an alpine leaf beetle: Divergence within *Oreina elongata* spans several ice ages " Molecular Phylogenetics and Evolution.

Bremer, K. (1988). "The limits of amino acid sequence data in angiosperm phylogenetic reconstruction." Evolution **42**: 795-803.

Brochmann, C., T. M. Gabrielsen, et al. (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." Proceedings of the National Academy of Sciences of the United States of America **91**: 6491-6495.

Buerki, S., F. Forest, et al. (in press). "An evaluation of new parsimony-based versus parametric inference methods in biogeography: a case study using the globally distributed plant family Sapindaceae." Journal of Biogeography.

de Queiroz, K. (2005). "Ernst Mayr and the modern concept of species." Proceedings of the National Academy of Sciences of the United States of America **102**: 6600-6607.

656 Després, L. and N. Jaeger (1999). "Evolution of oviposition strategies and speciation in the
657 globeflower flies *Chiastocheta* spp. (Anthomyiidae)." Journal of Evolutionary Biology
658 **12**: 822-831.
659 Després, L., E. Pettex, et al. (2002). "Speciation in the globeflower fly *Chiastocheta* spp.
660 (Diptera: Anthomyiidae) in relation to host plant species, biogeography, and
661 morphology." Molecular Phylogenetics and Evolution **22**(2): 258-68.
662 Drummond, A. J., S. Y. W. Ho, et al. (2006). "Relaxed phylogenetics and dating with
663 confidence." PLoS Biology **4**(5): e88.
664 Drummond, A. J. and A. Rambaut (2007). "BEAST: Bayesian evolutionary analysis by
665 sampling trees." BMC Evolutionary Biology **7**: 214.
666 Dufaÿ, M. and M.-C. Anstett (2003). "Conflicts between plants and pollinators that reproduce
667 within inflorescences: evolutionary variations on a theme." OIKOS **100**: 3-14.
668 Ehlers, J. and P. L. Gibbard (2004). Quaternary glaciations: extent and chronology.
669 Developments in Quaternary Science. Amsterdam; San Diego, Elsevier.
670 Espíndola, A., S. Buerki, et al. (2010). "New insights into the phylogeny and biogeography of
671 *Arum* L. (Araceae): unravelling its evolutionary history." Botanical Journal of the
672 Linnean Society **163**: 14-32.
673 Espíndola, A., L. Pellissier, et al. (in press). "Variation in the proportion of flower visitors of
674 *Arum maculatum* across its distributional range in relation with community-based
675 climatic niche analysis." OIKOS.
676 Gallet, C., S. Ibanez, et al. (2007). "Plant chemical defense induced by a seed-eating
677 pollinator mutualist." Journal of Chemical Ecology **33**(11): 2078-89.
678 Gelman, A. and D. B. Rubin (1992). "Inference from iterative simulation using multiple
679 sequences." Statistical Science **7**: 457-472.
680 Hall, T. A. (1999). "BioEdit: a user-friendly biological sequence alignment editor and analysis
681 program for Windows 95/98/NT." Nucleic Acids Symposium Series **41**: 95-98.
682 Hemborg, A. M. and L. Després (1999). "Oviposition by mutualistic seed-parasitic pollinators
683 and its effects on annual fitness of single- and multi-flowered host plants." Oecologia
684 **120**(3): 427-436.
685 Hennig, W., Ed. (1976). Anthomyiidae. Die Fliegen der Palaearktischen Region. Stuttgart, E.
686 Schweizerbart.
687 Ibanez, S. and L. Després (2009). "Ecological conditions promoting plant specialization on a
688 seed-eating pollinator differ from those stabilizing the interaction." Evolutionary
689 Ecology Research **11**: 921-934.
690 Ibanez, S., G. Dujardin, et al. (2009). "Stability of floral specialization in *Trollius europaeus* in
691 contrasting ecological environments." Journal of Evolutionary Biology **22**(6): 1183-
692 1192.

693 Ibanez, S., C. Gallet, et al. (2009). "Plant chemical defence: a partner control mechanism
694 stabilising plant--seed-eating pollinator mutualisms." BMC Evolutionary Biology **9**:
695 261.
696 Jaeger, N. and L. Després (1998). "Obligate mutualism between *Trollius europaeus* and its
697 seed-parasite pollinators *Chiastocheta* flies in the Alps." Comptes Rendus De
698 L'Académie Des Sciences Serie III-Sciences De La Vie-Life Sciences **321**(9): 789-
699 796.
700 Katoh, K. and H. Toh (2008). "Recent developments in the MAFFT multiple sequence
701 alignment program." Briefings in Bioinformatics **9**(4): 286-298.
702 Mallet, J. (2010). "Group selection and the development of the biological species concept."
703 Philosophical Transactions of the Royal Society of London Series B-Biological
704 Sciences **365**(1547): 1853-1863.
705 Mayr, E. (1942). Systematics and the origin of species, from the viewpoint of a zoologist.
706 Cambridge, Harvard University Press.
707 Meusel, H., E. Jäger, et al. (1965). Vergleichende Chorologie der zentraleuropäischen Flora.
708 Bd.1 (Text und Karten). Jena, Fischer.
709 Mishler, B. D. and E. C. Theriot (2000). The Phylogenetic Species Concept (*sensu* Mishler
710 and Theriot): monophyly, apomorphy, and Phylogenetic Species Concepts. Species
711 concepts and phylogenetic theory. Q. D. Wheeler and R. Meier. New York, Columbia
712 University Press**:** 44-54.
713 Nixon, K. C. (1999). "The parsimony ratchet, a new method for rapid parsimony analysis."
714 Cladistics **15**: 407-414.
715 Nylander, J. A. A. (2004). Mraic.pl. Uppsala, Sweden, Distributed by the author.
716 Papadopoulou, A., I. Anastasiou, et al. (2010). "Revisiting the insect mitochondrial molecular
717 clock: the mid-Aegean trench calibration." Molecular Biology and Evolution **27**(7):
718 1659-1672.
719 Pellmyr, O. (1989). "The cost of mutualism - interactions between *Trollius europaeus* and its
720 pollinating parasites." Oecologia **78**(1): 53-59.
721 Pellmyr, O. (1992). "The phylogeny of a mutualism: evolution and coadaptation between
722 *Trollius* and its seed-parasitic pollinators." Biological Journal of the Linnean Society
723 **47**: 337-365.
724 Pompanon, F., E. Pettex, et al. (2006). "Patterns of resource exploitation in four coexisting
725 globeflower fly species (*Chiastocheta* sp.)." Acta Oecologica **29**(2): 233-240.
726 Rambaut, A. and A. J. Drummond (2004). Tracer v1.4. Edinburgh.
727 R Development Core Team (2009). R: A Language and Environment for Statistical
728 Computing.

729 Ree, R. H., B. R. Moore, et al. (2005). "A likelihood framework for inferring the evolution of
730 geographic range on phylogenetic trees." Evolution **59**: 2299-2311.
731 Ree, R. H. and I. Sanmartín (2009). "Prospects and challenges for parametric models in
732 historical biogeographical inference." Journal of Biogeography **36**: 1211-1220.
733 Ree, R. H. and S. A. Smith (2008). "Maximum likelihood inference of geographic range
734 evolution by dispersal, local extinction, and cladogenesis." Systematic Biology **57**: 4-
735 14.
736 Ridley, M. (1997). Evolution. Oxford, Oxford University Press.
737 Ronquist, F. (1997). "Dispersal-vicariance analysis: a new approach to the quantification of
738 historical biogeography." Systematic Biology **46**: 195-203.
739 Ronquist, F. and J. P. Huelsenbeck (2003). "MrBayes 3: Bayesian phylogenetic inference
740 under mixed models." Bioinformatics **19**(12): 1572-1574.
741 Ruse, M. (1969). "Definitions of species in biology." The British Journal for the Philosophy of
742 Science **20**(2): 97-119.
743 Sanmartín, I. and F. Ronquist (2004). "Southern Hemisphere biogeography inferred by
744 event-based models: Plant versus animal patterns." Systematic Biology **53**(2): 216-
745 243.
746 Sikes, D. S. and P. O. Lewis (2001). PAUPRat, Beta software, version 1: PAUP
747 implementation of the parsimony ratchet, Distributed by the authors.
748 Simmons, M. P. and H. Ochoterena (2000). "Gaps as characters in sequence-based
749 phylogenetic analyses." Systematic Biology **49**(2): 369-381.
750 Sorenson, M. D. and E. A. Franzosa (2007). Treerot, version 3. Boston MA, USA, Distributed
751 by the authors.
752 Stamatakis, A. (2006). "RAxML-VI-HPC: Maximum Likelihood-based phylogenetic analyses
753 with thousands of taxa and mixed models." Bioinformatics **22**(21): 2688-2690.
754 Swofford, D. L. (2003). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other
755 Methods). Version 4. Sunderland, Massachusetts, Sinauer.
756 Tamura, K., J. Dudley, et al. (2007). "MEGA4: Molecular Evolutionary Genetics Analysis
757 (MEGA) software version 4.0." Molecular Biology and Evolution **24**: 1596-1599.
758 Van Valen, L. (1976). "Ecological species, multispecies, and oaks." Taxon **25**: 233-239.
759 Wahlberg, N. (2006). "That awkward age for butterflies: Insights from the age of the butterfly
760 subfamily Nymphalinae (Lepidoptera : Nymphalidae)." Systematic Biology **55**: 703-
761 714.
762 Zakharov, E. V., M. S. Caterino, et al. (2004). "Molecular phylogeny, historical biogeography,
763 and divergence time estimates for swallowtail butterflies of the genus *Papilio*
764 (Lepidoptera : Papilionidae)." Systematic Biology **53**(2): 193-215.
765
766

767 Table 1 – Regions, primer sequences, annealing temperatures and references used
768 for PCR amplification.

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	52	Després and Jaeger, 2000
COII	tRNAleu-3023	GAT TAG TGC AAT GGA TTT AGC TC	52	Després and Jaeger, 2001
	COII-3683	CCR CAA ATT TCT GAA CAT TGA CC	52	Després and Jaeger, 2002
D-loop	TM-N-193	TGG GGT ATG AAC CCA GTA GC	55	Simon et al. 1994
	SR-J-14612	AGG GTA TCT AAT CCT AGT TT	55	Simon et al. 1994
ITS2	ITS2DF	CCT GGT TAG TTT CTT TTC CTC CGC T	56	Depaquit et al., 2000
	ITS2DR	CGC AGC TAA CTG TGT GAA ATC	56	Depaquit et al., 2000

769

770 Table 2 – Number and percentage of constant (C), variable (V), parsimony informative (PI)
771 and total base pairs, for the four different amplified regions. Analysis of the D-loop did not
772 allow overlap between forward and reverse cycle sequencing (values are presented for each
773 primer/direction, independently). Percentages are shown between parentheses.

Mitochondrial	C	V	PI	Total
COI	489 (74.43)	168 (25.57)	77 (25.57)	657
COII	268 (55.95)	211 (44.05)	39 (8.14)	479
D-loop IF	277 (78.47)	76 (21.53)	49 (13.88)	353
D-loop IIR	577 (79.80)	238 (29.02)	155 (19.02)	815
Total				*2304*
Nuclear				
ITS 2	425 (96.37)	16 (3.63)	11 (2.49)	441

774

Table 3 – Contingency tables of assignments of samples to a) mitochondrial *vs.* nuclear clades; b) nuclear *vs.* typological/morphological species; c) mitochondrial *vs.* typological/morphological species. Names correspond to clades following Figure 2. In a), NA indicates samples that have been only analyzed using the mitochondrial approach.

a)

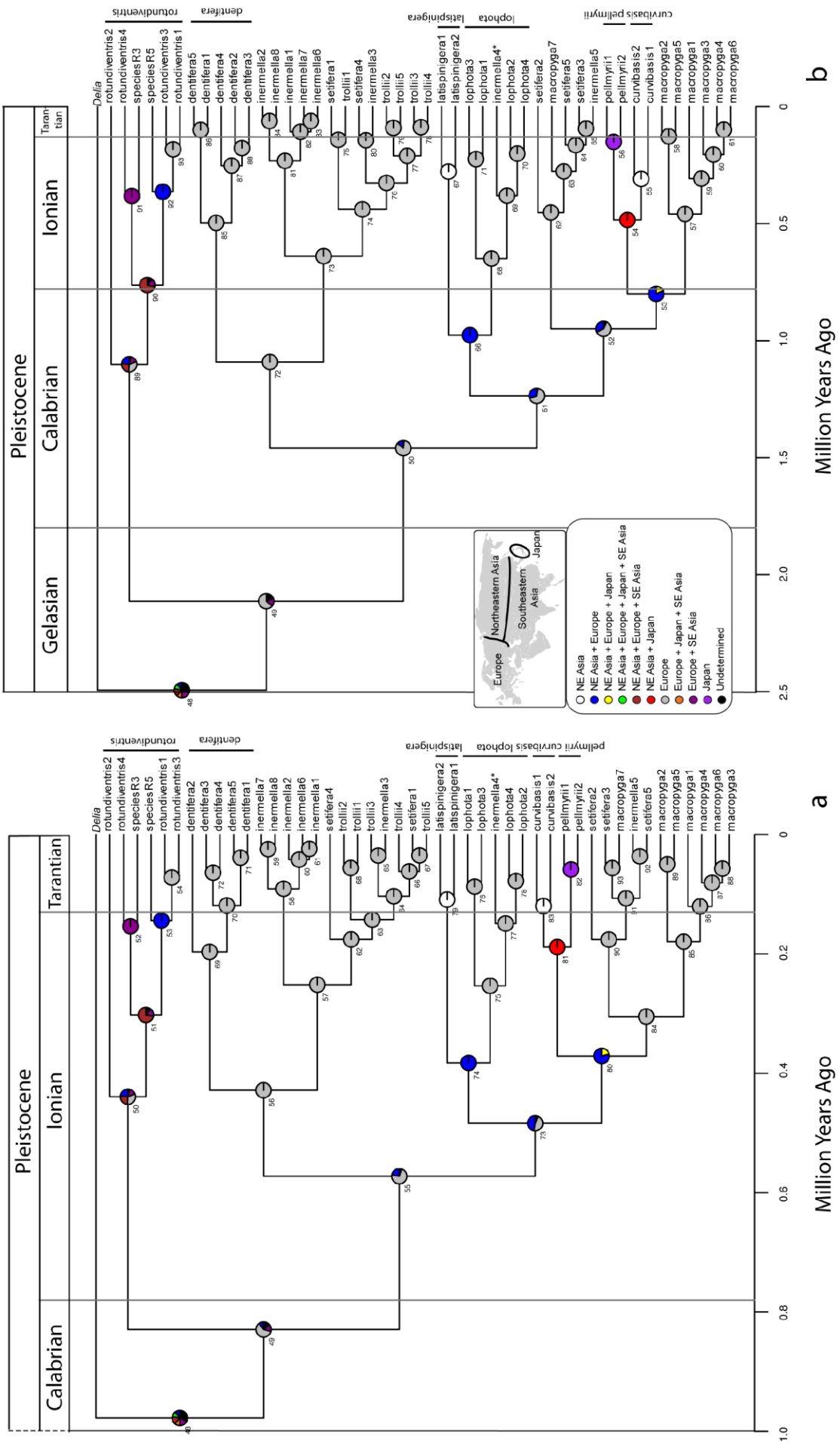
Mitochondrial clade	Nuclear clade: rotundiventris	trollii	lophota	setifera	nI	NA	**Total**
rotundiventris	40					49	89
dentifera					13	25	38
lophota			36			11	47
clade mtI		34		14	38	52	138
clade mtII			2	18	40	27	87
Total	40	33	38	32	92	164	**399**

b)

Morphospecies	Nuclear clade: rotundiventris	trollii	lophota	setifera	nI	**Total**
C. rotundiventris	40					40
C. dentifera					14	14
C. lophota			37		1	38
C. inermella				2	38	40
C. trollii		30			2	32
C. macropyga					28	28
C. setifera		4	1	30	8	43
Total	40	34	38	32	91	**237**

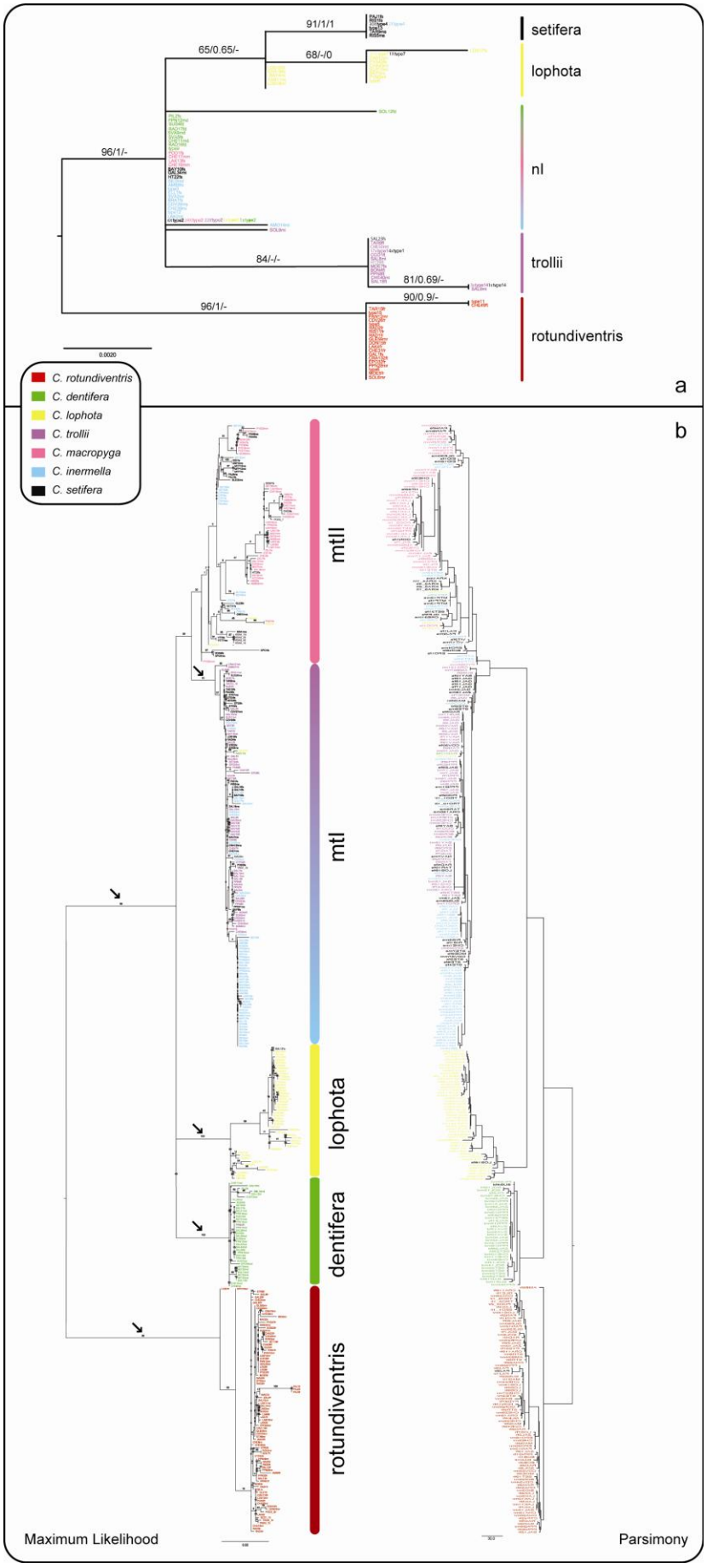
c)

Morphospecies	Mitochondrial clade: rotundiventris	dentifera	lophota	mtI	mtII	**Total**
C. rotundiventris	72					72
C. dentifera		37		1		38
C. lophota			45	1	3	49
C. inermella				53	14	67
C. trollii		1		53		54
C. macropyga					41	41
C. setifera			1	27	24	52
larvae	17		1	3	5	26
Total	89	38	47	138	87	**399**


b
Pleistocene
Gelasian
Calabrian
Ionian
Tarantian
Million Years Ago
2.5
2.0
1.5
1.0
0.5
0
rotundiventris
dentifera
latispinigera
lophota
curvibasis pellmyrii
Delia
rotundiventris2
rotundiventris4
speciesR3
speciesR5
rotundiventris3
rotundiventris1
dentifera5
dentifera1
dentifera4
dentifera2
dentifera3
inermella2
inermella8
inermella1
inermella7
inermella6
setifera1
trollii1
setifera4
inermella3
trollii2
trollii5
trollii3
trollii4
latispinigera1
latispinigera2
lophota3
lophota1
inermella4*
lophota2
lophota4
setifera2
macropyga7
setifera5
setifera3
inermella5
pellmyrii1
pellmyrii2
curvibasis2
curvibasis1
macropyga2
macropyga5
macropyga1
macropyga3
macropyga4
macropyga6
Europe
Northeastern Asia
Southeastern Asia
Japan
NE Asia
NE Asia + Europe
NE Asia + Europe + Japan
NE Asia + Europe + Japan + SE Asia
NE Asia + Europe + SE Asia
NE Asia + Japan
Europe
Europe + Japan + SE Asia
Europe + SE Asia
Japan
Undetermined
a
Pleistocene
Calabrian
Ionian
Tarantian
Million Years Ago
1.0
0.8
0.6
0.4
0.2
0
rotundiventris
dentifera
latispinigera
lophota
pellmyrii curvibasis
Delia
rotundiventris2
rotundiventris4
speciesR3
speciesR5
rotundiventris1
rotundiventris3
dentifera2
dentifera3
dentifera4
dentifera5
dentifera1
inermella7
inermella8
inermella2
inermella6
inermella1
setifera4
trollii2
trollii1
trollii3
inermella3
trollii4
setifera1
trollii5
latispinigera2
latispinigera1
lophota1
lophota3
inermella4*
lophota4
lophota2
curvibasis1
curvibasis2
pellmyrii1
pellmyrii2
setifera2
setifera3
macropyga7
inermella5
setifera5
macropyga2
macropyga5
macropyga1
macropyga4
macropyga6
macropyga3

787

788 Figure 1 – Dating and biogeographic reconstruction of twelve species within the genus
789 *Chiastocheta*. Portions in pies indicate ancestral areas assigned by the Dispersion-
790 Extinction-Cladogenesis (DEC) approach. Colors indicate ancestral areas (see figure
791 legend) and map shows biogeographic areas. a – inference considering a substitution rate of
792 2.28%; b - inference considering a substitution rate of 0.9%. Vertical bars delimitate main
793 geological periods. "*" indicates a putative wrongly identified sample in previous studies.
794


91/1/1
65/0.65/-
68/-/0
96/1/-
84/-/-
81/0.69/-
90/0.9/-
96/1/-
setifera
lophota
nl
trollii
rotundiventris
a
C. rotundiventris
C. dentifera
C. lophota
C. trollii
C. macropyga
C. inermella
C. setifera
b
mtII
mtI
lophota
dentifera
rotundiventris
Maximum Likelihood
Parsimony

795

796 Figure 2 – Phylogenetic inferences based on (a) nuclear and (b) mitochondrial data and
797 rooted considering relative taxon positions suggested in Figure 1 and Després et al. (2002).
798 Colors of tips indicate morphological species (refer to legend in figure). In a, node supports
799 are shown on branches (ML/Bayesian/MP); in b, main supported clades are shown with an
800 arrow.
801

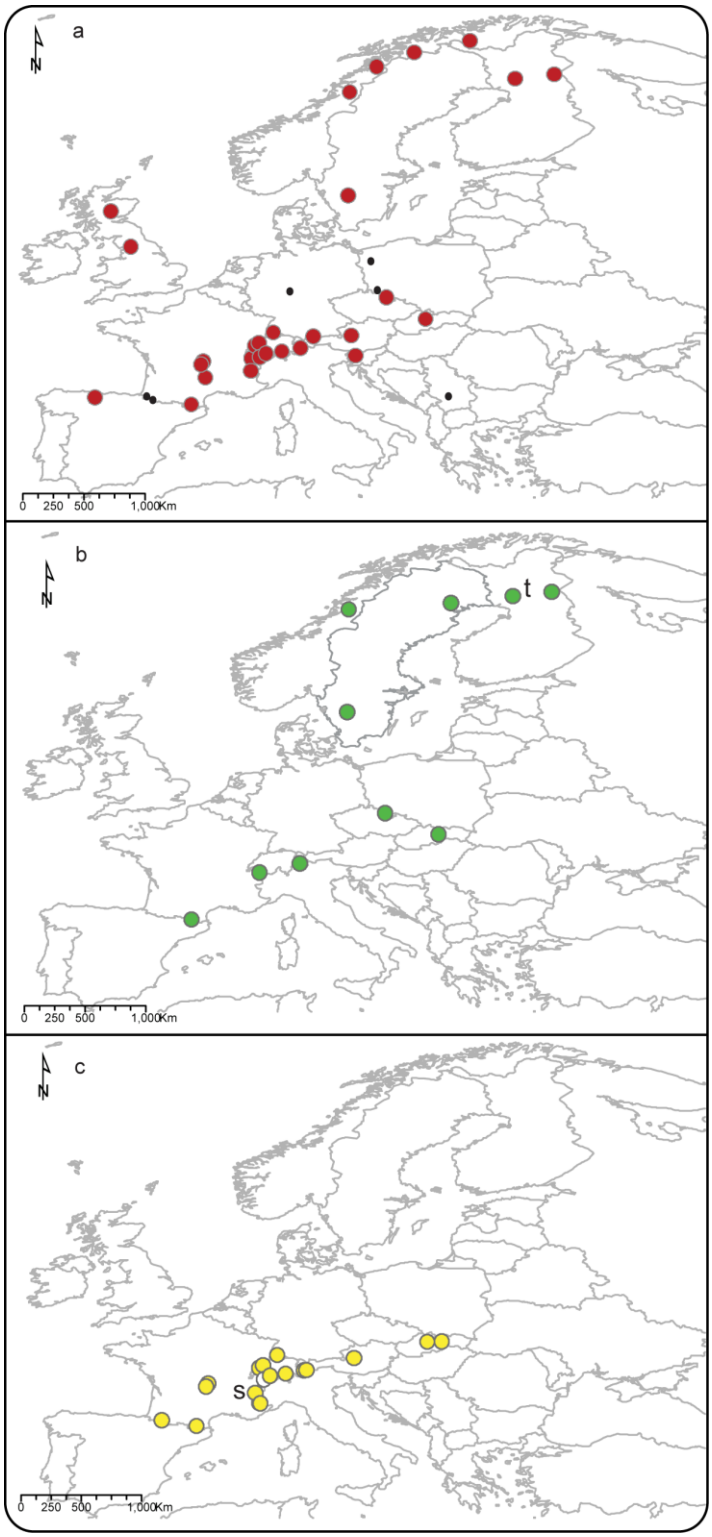


Figure 3 – Distribution of mitochondrial clades for which congruence between morphological characters and molecular data was observed. a – *C. rotundiventris* , b – *C. dentifera*, c- *C. lophota*. Letters in b and c represent other exceptional morphospecies identified in the clades; t: *C. trollii*; s: *C. setifera.* Black dots in figure (a) indicate sites for which larvae samples fell in the mitochondrial clade.

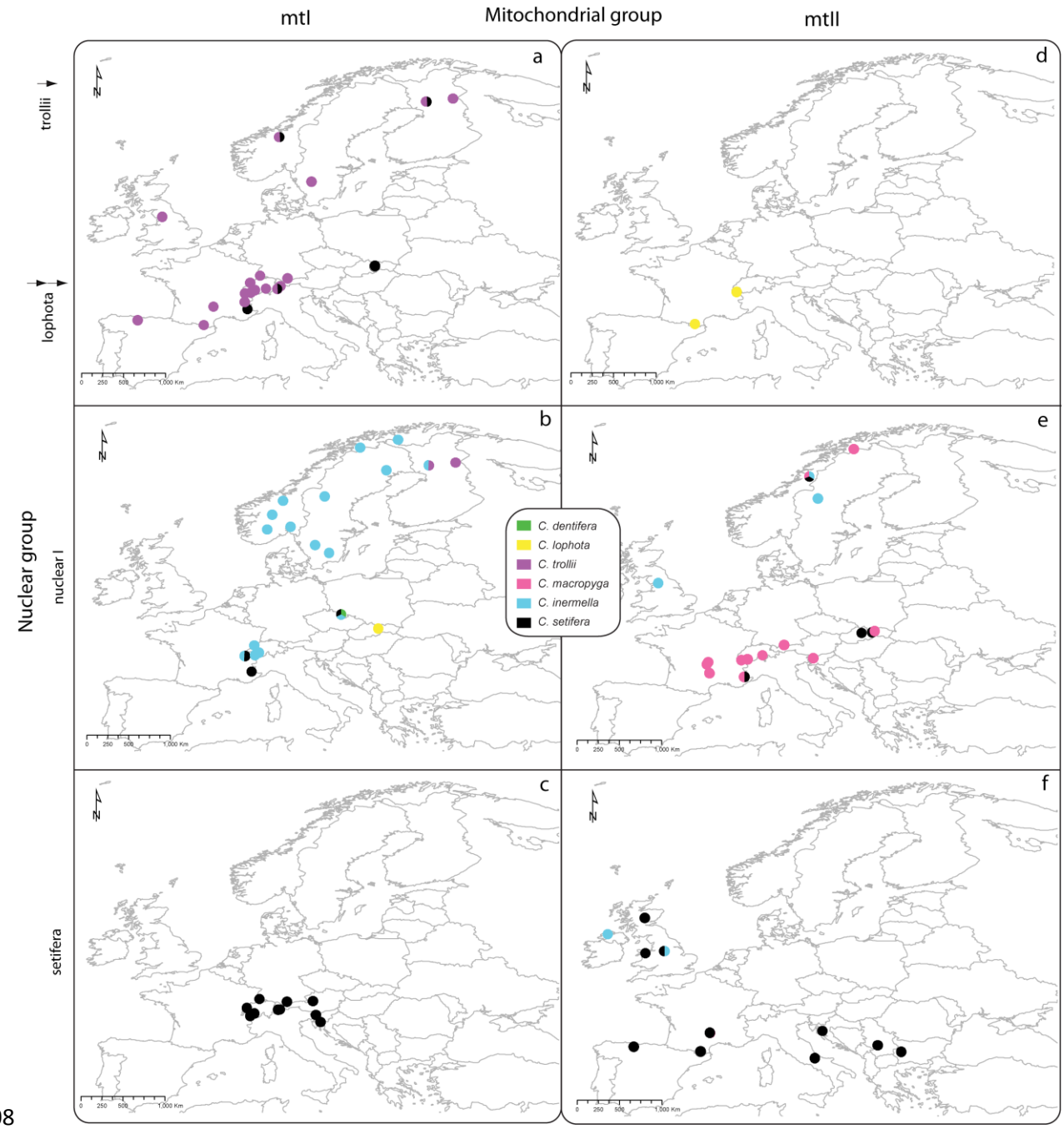


Figure 4 – Distribution and morphospecies composition of samples included in mitochondrial groups “mtI” (a-c) and “mtII” (d-f), in respect to nuclear groups in which they were included (a: “trollii”; b and e: “nI”; c and f: “setifera”; d: “lophota”). Colors indicate morphospecies; for explanations refer to the figure legend.

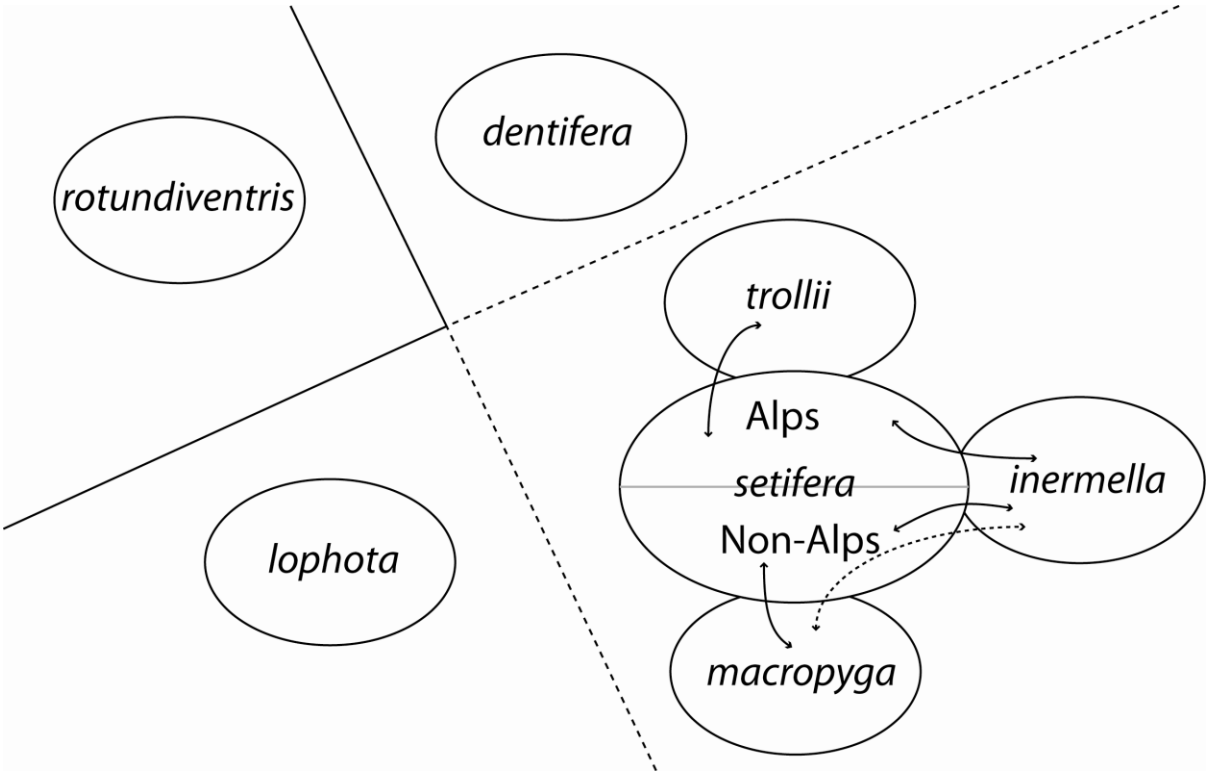


Figure 5 – Hybridization scheme in *Chiastocheta* flies pollinating *T. europaeus.* Lines indicate complete isolation; dashed lines indicate important isolation with possible hybridization; arrows indicate hybridizations; dashed arrows show possible indirect genome transfer.

822 Supplementary Material 1 – Samples included in the dating and biogeographic analyses

Name	Sample_ID	Origin	GenBank	Species
curvibasis1	curviba2	GenBank	AH010127.1	*C. curvibasis*
curvibasis2	curvSkr3	GenBank	AH010133.1	*C. curvibasis*
Delia	Delia	GenBank	AH010111.1	*Delia brassicae*
dentifera1	CHE10md	CH Chemin	This study	*C. dentifera*
dentifera2	denti3	GenBank	AH010114.1	*C. dentifera*
dentifera3	RAD16fd	Radkow	This study	*C. dentifera*
dentifera4	SET25fd	Seterasen	This study	*C. dentifera*
dentifera5	SOL14md	Solberga	This study	*C. dentifera*
inermella1	AMO8fi	Amot	This study	*C. inermella*
inermella2	BEI15fi	Beistohlen	This study	*C. inermella*
inermella3	CHE5fi	CH Chemin	This study	*C. inermella*
inermella4	inerm3	GenBank	AH010117.1	*C. inermella*
inermella5	inerm4	GenBank	AH010118.1	*C. inermella*
inermella6	inerm5	GenBank	AH010764.1	*C. inermella*
inermella7	inerm6	GenBank	AH010765.1	*C. inermella*
inermella8	SAL23fi	Salla	This study	*C. inermella*
latispinigera1	latispSKr2	GenBank	AH010132.1	*C. latispinigera*
latispinigera2	latisSKz4	GenBank	AH010135.1	*C. latispinigera*
lophota1	DON3fl	Donovaly	This study	*C. lophota*
lophota2	HT14fl	Haute Tinee 1	This study	*C. lophota*
lophota3	LOS19ml	Loser	This study	*C. lophota*
lophota4	MOE8ml	Moerlimatt	This study	*C. lophota*
macropyga1	CHE20mm	CH Chemin	This study	*C. macropyga*
macropyga2	LAK13fm	Laktatjakka	This study	*C. macropyga*
macropyga3	macro2	GenBank	AH010120.1	*C. macropyga*
macropyga4	macro4	GenBank	AH010766.1	*C. macropyga*
macropyga5	macrop3	GenBank	AH010768.1	*C. macropyga*
macropyga6	POD1fm	Podlesok	This study	*C. macropyga*
macropyga7	PYD63fm	Puy de Dome (FR)	This study	*C. macropyga*
pellmyrii1	pellmy1	GenBank	AH010769.1	*C. pellmyrii*
pellmyrii2	pellmy2	GenBank	AH010767.1	*C. pellmyrii*
rotundiventris1	CHE25mr	CH Chemin	This study	*C. rotundiventris*
rotundiventris2	LOS7fr	Loser	This study	*C. rotundiventris*
rotundiventris3	PAN12mr	Puerto de Panderruedas	This study	*C. rotundiventris*
rotundiventris4	ZAL10mr	Zali Log	This study	*C. rotundiventris*
setifera1	BAY9fs	Bayasse	This study	*C. setifera*
setifera2	EID1fs	Eidda Pastures	This study	*C. setifera*
setifera3	MTP1fs	Monte Pizi	This study	*C. setifera*
setifera4	setifera	GenBank	AH010123.1	*C. setifera*
setifera5	VIT17ms	Vitosha	This study	*C. setifera*
speciesR3	spR3	GenBank	AH010125.1	species R3
speciesR5	spR5	GenBank	AH010131.1	species R5
trollii1	EPO11mt	Esposouille	This study	*C. trollii*
trollii2	GAL36mt	Col du Galibier (FR)	This study	*C. trollii*
trollii3	MOE7ft	Moerlimatt	This study	*C. trollii*
trollii4	NAV17ft	Naverdal	This study	*C. trollii*
trollii5	trollii	GenBank	AH010122.1	*C. trollii*

Supplementary Material 2 – *Q* matrix used for biogeographic reconstructions. A: Central Asia; B: Europe; C: Japan; D: Southeastern Asia. Values indicate probability of dispersion between areas.

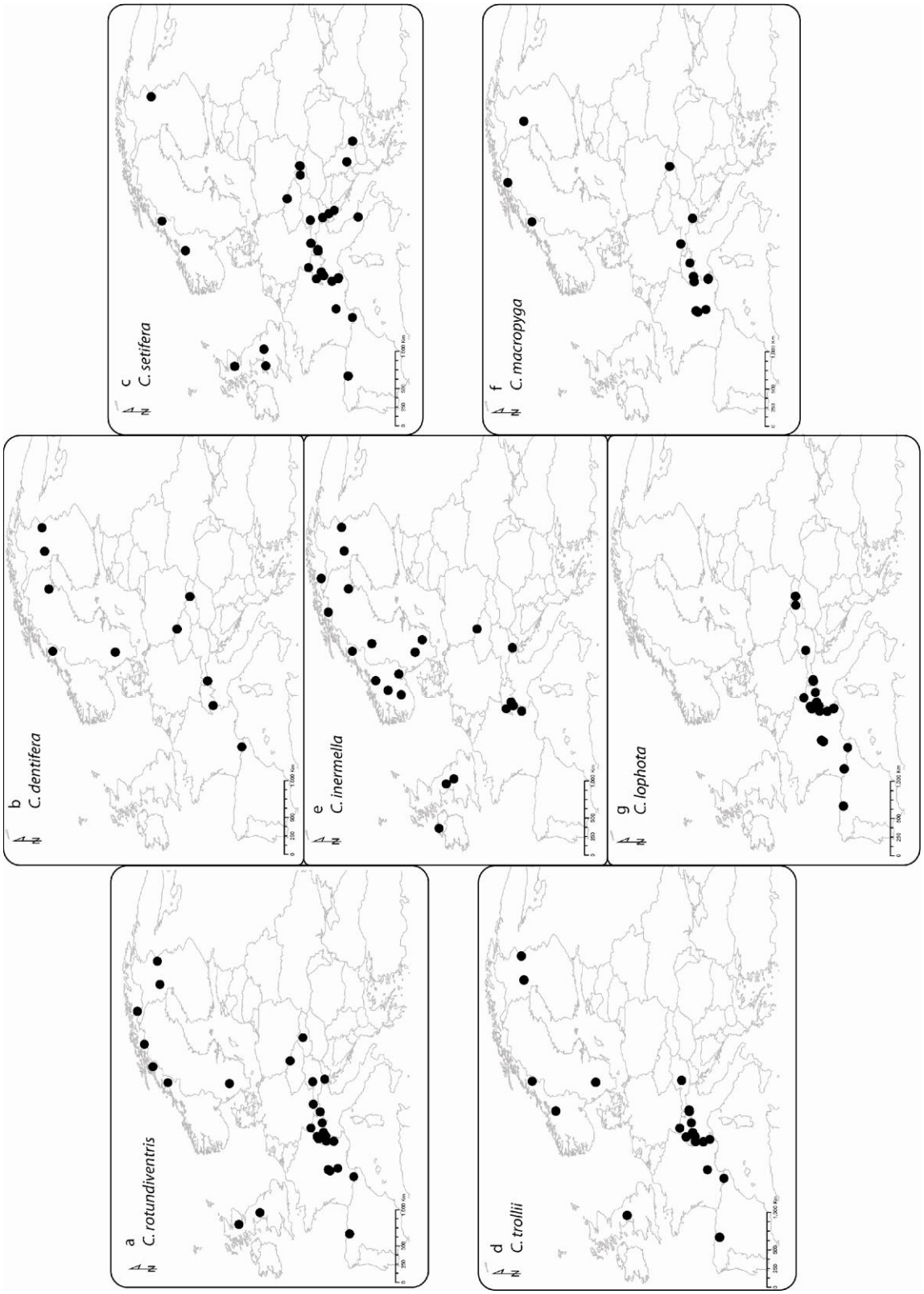
	A	**B**	**C**	**D**
A	-	1	1	1
B	1	-	0.01	0.01
C	1	0.01	-	1
D	1	0.01	1	-

828 Supplementary Material 3 – Sampled locations and number of mosphospecies used in molecular analyses.

Location	Longitude	Latitude	Altitude (m)	Collection date	dentifera	inermella	lophota	macropyga	rotundiventris	setifera	trollii	larvae	Total
Ambri	8.702920	46.506800	1000	30.5.08	0	0	2	4	2	0	1	0	9
Amot	8.42346	59.62199	481	18.06.07	0	4	0	0	0	0	0	0	4
Bayasse	6.740670	44.308140	N/A	9.06.07	0	0	2	2	0	2	2	0	8
Beistohlen	8.95473	61.20761	729	19.06.07	0	5	0	0	0	0	0	0	5
Bidjovagge	22.478080	69.297780	613	1.8.08	0	1	0	0	2	0	0	0	3
Bobolice	16.59895	53.94660	110	19.6.08	0	0	0	0	0	0	0	2	2
Col de Bonnecombe	3.11410	44.57557	1335	21.05.07	0	0	0	3	1	1	3	0	8
Braas	15.06817	57.09309	209	15.06.07	0	1	0	0	0	0	0	0	1
Col de la Colombière (FR)	6.469722	45.98722	1600	14.06.06	0	0	1	1	1	0	1	0	4
CH Creux du Van	6.741193	46.93526	N/A	29.06.06	0	1	2	0	2	2	0	0	7
CH Chasseral	7.021299	47.12569	N/A	29.06.06	0	0	5	0	5	0	1	0	11
CH Chemin	7.089778	46.08993	N/A	26.05.06	4	4	1	6	7	3	5	0	30
CH Crans-Montana	7.538896	46.34650	N/A	25.05.06	0	1	1	0	3	1	1	0	7
Cressbrook Dale	-1.740410	53.267240	246	10.6.08	0	1	0	0	0	1	0	0	2
Colt Park	-2.352470	54.193650	380	10.6.08	0	4	0	0	1	0	1	0	6
Donovaly	19.230680	48.889220	1345	21.6.08	0	0	3	0	2	1	0	0	6
Eidda Pastures	-3.741900	53.037200	234	9.6.08	0	0	0	0	0	2	0	0	2
Ellingsrudelva	10.91844	59.91771	161	18.06.07	0	1	0	0	0	0	0	0	1
Esposouille	2.094500	42.623410	1521	3.6.08	1	0	1	0	2	2	3	0	9
Froson	14.60268	63.18205	424	20.06.07	0	2	0	0	0	0	0	0	2
Col du Galibier (FR)	6.438611	45.08528	2000	14.06.06	0	3	3	0	2	3	7	0	18
Glen Fender	-3.794850	56.781380	345	14.6.08	0	0	0	0	4	2	0	0	6
Haute Tinee 1	6.818710	44.296170	2020	4.06.07	0	0	3	0	0	0	0	0	3
Haute Tinee 2	6.855810	44.284260	1770	4.06.07	0	0	0	2	0	1	0	0	3
Krasno Polje	14.97271	44.808690	1487	9.7.08	0	0	0	0	0	1	0	4	5

Laktatjakka	18.40674	68.42931	453	22.06.07	0	2	0	10	7	0	0	0	19
Lough Fern	-7.711300	55.065690	26	13.6.08	0	1	0	0	0	0	0	0	1
Loser	13.78485	47.66052	1598	04.06.07	0	0	5	0	4	1	1	0	11
Col long de Magnabaigt	-0.43646	42.87060	1615	23.7.08	0	0	0	0	0	0	0	4	4
Moerlimatt	8.07760	47.90597	939	10.06.07	0	0	1	0	1	1	1	0	4
Monte Pizi	14.167140	41.915240	1546	28.5.08	0	0	0	0	0	3	0	0	3
Naverdal	10.13002	62.70417	480	19.06.07	0	10	0	0	0	1	6	0	17
Pajino Preslo	20.819700	43.277990	1802	8.7.08	0	0	0	0	0	2	0	3	5
Puerto de Panderruedas	-4.972230	43.127430	1227	4.6.08	0	0	2	0	4	2	1	0	9
Pila	20.294490	48.900170	972	21.6.08	2	0	3	0	0	1	0	0	6
Podlesok	20.351900	48.949620	579	21.6.08	0	0	0	1	0	1	0	2	4
petit papa noel	25.79386	66.51647	72	23.06.07	6	4	0	1	4	2	4	0	21
Puy de Dome (FR)	2.963333	45.77222	1460	15.06.06	0	0	2	2	1	0	0	0	5
Puy de Sancy (FR)	2.809722	45.53500	1520	16.06.06	0	0	6	4	1	0	0	0	11
Radkow	16.353210	50.468660	712	20.6.08	2	2	0	0	2	2	0	5	13
Risnjak - Snjeznik	14.584940	45.438710	1466	5.06.08	0	0	0	0	0	2	0	0	2
Salla	28.65427	66.83020	194	23.06.07	7	5	0	0	3	1	6	0	22
Sede de Pan	-0.486510	43.039490	1556	22.7.08	0	0	2	0	0	0	0	2	4
Seterasen	13.67744	65.53432	285	21.06.07	6	7	0	1	1	1	2	0	18
Solberga	13.56116	57.95194	239	15.06.07	4	4	0	0	3	0	1	0	12
Steingaden	11.01296	47.59529	1158	30.05.07	0	0	0	1	1	6	0	0	8
Straumen	15.64921	67.38440	75	21.06.07	0	0	0	0	3	0	0	0	3
CH Susch	10.074725	46.747277	N/A	05.06.06	2	0	2	0	1	1	3	0	9
Svartla	21.22062	65.99583	36	22.06.07	4	3	0	0	0	0	0	0	7
CH Tarasp	10.250556	46.777299	N/A	05.06.06	0	0	2	0	0	2	2	0	6
Trollblumenwiese	11.04118	51.68314	488	16.6.08	0	0	0	0	0	0	0	4	4
Vitosha	42.590320	23.293420	1779	3.7.08	0	0	0	0	0	2	0	0	2
Zali Log	14.110800	46.203420	533	17.5.08	0	1	0	3	2	1	0	0	7

Supplementary Material 4 –Distribution ranges of morphologically identified (typological) species. a- *C. rotundiventris,* b- *C. dentifera,* c- *C. setifera,* d- *C. trollii,* e- *C. inermella,* f- *C. macropyga,* g- *C. lophota.*



Morphospecies and their distribution ranges

Our study confirms some indications provided by Pellmyr (1992), which proposed distribution ranges for the different *Chiastocheta* species associated with *Trollius europaeus*. While he based his conclusions on egg identifications, we studied adult distributions, and even if some results are compatible with those previously presented, other strongly differ.

Different patterns appeared depending on the morphological group analyzed: while some species are widely distributed and cover the whole distribution range (first case), others present a fragmented distribution or are restricted to some regions (second case).

In the first case, we found samples belonging to *C. rotundiventris* and to some extent *C. dentifera, C. trollii*, *C. setifera* and *C. inermella*. The first species (Figure a) presents in our study a similar distribution range than that depicted by Pellmyr in his revision of the *Trollius-Chiastocheta* interaction (Pellmyr, 1992). We here also confirm the presence of this taxon in the Eastern Alps and the Tatra region, a zone designed by the last surveys as putative for the presence of flies. This geographic constancy could indicate that this species is also the most reliable for pollination.

C. dentifera (Fig b) is also present in all regions, but with rarer occurrences than *C. rotundiventris* (ten *vs.* 25 locations, respectively). The distribution range presented by our study shows some congruence with that obtained by Pellmyr (1992), since we did not find any *C. dentifera* in the British Isles, but we did observe them in the Pyrenees. Differences in the presence and absence of *C. dentifera* in the British Isles in Pellmyr's and our study, respectively, could be either due to a recent change in the distribution range of this species in that region –the last records in the zone date back to 1960 (NBN, 2010) and correspond exactly to those presented in Pellmyr's work– or to a random lack of this species in our sampling. This should be however easily tested by new surveys in the area.

C. setifera (Figure c) also occupies an important part of the sampled area, and is the only present in the Italian Peninsula and Balkan zone. This last point would indicate its ecological importance in these regions, since the lack of other species would design *C. setifera* as the only species available to pollinate flowers. Moreover, our results also extend the distribution

range of this species to Scandinavia, a region from which it was completely lacking during the last surveys.

Results on the distribution range of *C. trollii* (Figure d) correspond well to distributions obtained based on egg morphology by Pellmyr (1992). However, we additionally show the presence of this species in the Pyrenees and Cantabria.

The last species displaying a large distribution is *C. inermella* (Figure e), which appeared in our survey as clearly present in Scandinavia, but also in the Alps, the Tatra and the British Isles. This difference to Pellmyr (1992) is substantial and would tend to reject a previous hypothesis of a vicariant identity between this typological species and *C. lophota* (its supposed Southern vicariant; Pellmyr, 1992).

Only *C. macropyga* (Figure f) appeared to present a fragmented distribution range. The species is present in the Alpine chain, the French Massif Central and to some extent the Tatra region. However, there are no records of the species between these zones and Northern Scandinavia, the other region where it has been observed. This distribution confirms the view proposed by Pellmyr (1992).

Finally, the only species appearing restricted to only some regions is *C. lophota* (Figure g), which is mentioned by Pellmyr (1992) as the Southern vicariant of *C. inermella* (despite our results contradict this view). Our survey shows that the species is thus only present in the Southern mountainous chains of Europe.

References

Pellmyr, O. (1992) The phylogeny of a mutualism: Evolution and coadaptation between *Trollius* and its seed-parasitic pollinators. *Biological Journal of the Linnean Society*, **47**, 337-365

The National Biodiversity Network (NBN), UK. 2010. http://www.nbn.org.uk/.

Chapter seven

No fossils, no answers?
Inferring postglacial genetic consequences in cold-adapted species when neither ancient DNA nor macroremains are available

Anahí Espíndola, Loïc Pellissier and Nadir Alvarez

Article type: Original article

Title: No fossils, no answers? Inferring postglacial genetic consequences in cold-adapted species when neither ancient DNA nor macroremains are available

Authors: Espíndola Anahí[1] *, Pellissier Loïc[2] and Alvarez Nadir[2]

[1]Laboratory of Evolutionary Entomology, Institute of Biology, University of Neuchâtel. Emile-Argand 11, 2009 Neuchâtel, Switzerland.

[2]Department of Ecology and Evolution, Biophore Building, University of Lausanne, 1015 Lausanne, Switzerland.

* Correspondence: Anahí Espíndola, Laboratory of Evolutionary Entomology, Institute of Biology, University of Neuchâtel. Emile-Argand 11, 2009 Neuchâtel, Switzerland. Fax: +41 (0)32 718 3001. Email: maria.espindola@unine.ch

Running head: Postglacial contraction in cold-adapted taxa

ABSTRACT

Aim: Unlike temperate species, whose distribution ranges have expanded since the end of the Last Glacial Maximum (LGM), cold-adapted species have experienced substantial postglacial range contractions. To understand the dynamics of their contraction process, researchers have proposed to confront current phylogeographic patterns with past genetic and/or distribution data, inferred from ancient material. However, because fossils and/or macroremains are unavailable for most species, direct inferences on postglacial histories are only possible for a restricted number of taxa. In this study we present a novel approach to fill this technical gap and to open-up new possibilities for inferring the postglacial fate of organisms for which no ancient material is available.

Location: West-Palearctic.

Methods: Using *Trollius europaeus* (Ranunculaceae) as a case study, we sampled the whole European range of the plant and further inferred its current spatial genetic structure based on AFLP data. Simultaneously, we defined the Ecological Niche Model (ENM) of the species and used it to hindcast its LGM distribution. Based on this, we identified past suitable areas (suitability hotspots), which could have potentially hosted large numbers of individuals and specific genetic lineages at the LGM, and compared their distribution with the current spatial genetic structure of the plant. Because isolation of these hotspots might have driven lineage divergence at LGM, this comparative approach provides a probabilistic view of the distribution of past genetic diversities. We thus propose that the level of congruence between the number and spatial distribution of current genetic clusters and hindcasted LGM hotspots is a good proxy for informing on the genetic diversity loss induced by the important range contractions experienced after the LGM.

Results: The current genetic structure of the plant is spatially divided in four well-defined clusters, which correspond to four out of the five hindcasted LGM suitability hotspots. Assuming that lineages were not shared between hotspots, this discrepancy informs on a probable post-glacial loss of one genetic lineage in the species.

Main conclusions: Our approach provides a valuable probabilistic view on the genetic consequences of the last postglacial contractions in cold-adapted species, especially those for which other inferences are impossible because of the absence of ancient material.

Keywords

hindcasting, Ecological Niche Modelling, *Trollius europaeus*, spatial genetic structure, postglacial contraction, phylogeography, arctic-alpine species

INTRODUCTION

Species distributions are constrained both by physiological limitations related to abiotic factors (*e.g.*, temperature, moisture) and by the effect of biotic interactions in which they are involved (e.g., competition, parasitism; Pulliam, 2000). As a consequence, species are only able to occupy the subset of the available environment where their requirements are fulfilled, a concept referred to as the ecological niche by Hutchinson (1957). The spatial distribution of species is thus both mainly defined by their respective ecological niche, their dispersal characteristics and their colonization histories (Alvarez *et al.*, 2009). Therefore, any environmental change has the potential to induce large variations in the area occupied by a species, and to consequently influence other characteristics of the studied organism, as for instance, demographic and genetic parameters at the population level. If the distribution range of a species becomes fragmented and gene flow among populations is limited or absent, evolutionary processes (*e.g.*, vicariance, drift, selection) are expected to drive differences in adaptive and neutral genetic variation among isolated patches of populations (Hewitt, 1999). Hence, the genetic variation within a species is not only defined by intrinsic biological characteristics such as its reproductive strategy or its generation time, but also by extrinsic factors such as climate, topography and habitat availability (Avise, 2009).

Because of the interaction between orbital and tectonic forces, the Quaternary has been marked by cyclic climatic oscillations, defined by successive glacial and interglacial periods that have modified species' habitat availability through space and time (Hewitt, 1996). During the Last Glacial Maximum (LGM: 23 -18 thousand years ago -kya) most of northern Europe was covered by polar ice caps and glaciers filled internal valleys of southern European mountain ranges. Moreover, because of the large volume of water contained in the ice caps, sea levels lowered and new lands became exposed, especially in the English Channel and the Adriatic Sea (Ehlers & Gibbard, 2007). Consequently, biomes were displaced southwards and cold steppic environments could be found as far South as southern France, the Jura and southern Carpathians (Frenzel, 1968; Hewitt, 1999). While the main part of the continent presented suitable habitats for cold-adapted species (i.e., currently arctic, alpine or arctic-

alpine), southern regions harboured conditions more favourable to temperate organisms. This was mainly due to the proximity of the Mediterranean Sea that provided humidity and higher temperatures to surrounding lands, but also to the topography of southern mountains that offered a wide range of climatically protected habitats (Médail & Diadema, 2009; Stewart *et al.*, 2010). Demographic and distributional responses to climatic variations are expected to differ between temperate and cold-adapted species (reviewed in Stewart *et al.*, 2010): while temperate species are expected to be widespread during interglacial and restricted to refugia during glacial periods, cold-adapted species reflect the opposite demographic pattern.

For the last two decades, the biogeography of temperate species has been widely studied in the western Palearctic and phylogeographic patterns have been summarized into a few paradigms as the ones represented by the European beech *Fagus sylvatica,* the hedgehog *Erinaceus* and the bear *Ursus arctos* (e.g., Taberlet *et al.*, 1998; Hewitt, 1999; Schmitt, 2007). In contrast, probably because of more difficult sampling strategies associated to their arctic-alpine distributions, cold-adapted species received less attention and, compared to temperate species, only a few of their corresponding phylogeographic patterns were unravelled (Brochmann *et al.*, 2003; Schönswetter *et al.*, 2006; Ehrich *et al.*, 2007). Whereas investigating the spatial genetic structure of temperate organisms (currently showing an expanded distribution) provides information on the genetic consequences of postglacial expansion processes, studying the phylogeography of cold-adapted species, which are currently found under refugial conditions (*i.e.* at high altitude and/or latitude areas), allows scenario testing on the genetic consequences of distributional contractions. Understanding consequences of such contractions is not only important for historical reasons, it also provides information for predicting the fate of species expected to be negatively affected by the current and future global climate warming (Hampe & Petit, 2005).

Direct inference on the variation and temporal maintenance of the genetic diversity associated with previously expanded cold-adapted lineages is currently only possible when ancient DNA is available (Stewart *et al.*, 2010). For instance, Dalén *et al.* (2007) have shown that in the specific case of the polar fox, several ancient lineages – tracked back by the

analysis of ancient DNA – have nowadays become extinct and their respective alleles are absent from the current species' distribution. Even though such an approach brings a substantial contribution to our knowledge on the genetic consequences of climatic oscillations, it is restricted to the study of taxa for which ancient DNA or macroremains are accessible. Because there is a lack of available ancient DNA in some cold-adapted species, retracing the evolution of ancient lineages in organisms with current arctic or alpine distributions is thus limited to indirect inference.

In the present work, we propose an indirect technique to infer the fate of Quaternary lineages when ancient DNA or fossils are not available. By combining Ecological Niche Modelling (ENM) and further hindcasted climatic probabilities of occurrence, with an estimation of the current spatial genetic structure of a given organism, our approach allows inferring the probable contraction dynamics of species after the end of the European LGM. Based on a solid phylogeographic paradigm (Avise *et al.*, 1987; Avise, 2009), we suggest that hotspot areas showing high hindcasted probabilities of past suitability are correlated with a high probability of past occurrence of one or several endemic genotype(s) in that region. We thus aim to provide a probabilistic sight on the past genetic diversity of a species. Here we propose first to identify ENM-based *suitability hotspots*, defined as well-circumscribed regions showing high climatic suitability for a given species at the LGM. Second, based on a sampling covering a substantial portion of the current species distribution and applying genome fingerprinting techniques, we evaluate its present genetic structure. Finally, we combine these two approaches and test the following hypotheses: if the number and localization of LGM *suitability hotspots* correspond to the number and distribution of the current genetic lineages identified, it is likely that at least one sample per hotspot was preserved after the end of the last glacial period and contributed to the current genetic variation of the species; alternatively, if this was not the case, we can consider that the range contraction might have induced a possible loss of genetic diversity (Dalén *et al.*, 2007).

In order to test these hypotheses, we use the globeflower *Trollius europaeus* L. (Ranunculaceae) as a case study. This Eurasian arctic-alpine species is currently associated

with cold and humid habitats and covers the southern mountainous ranges of Europe as well as some northeastern European plains, extending up to northern Scandinavia and the Ural (Meusel *et al.*, 1965). As all arctic-alpine species, *T. europaeus* is considered to be currently in a retracted demographical situation and its distribution during the LGM is expected to have been more extensive than today (Stewart *et al.*, 2010). Based on a thorough sampling covering the whole European range of the plant distribution, we evaluate the amount of past genetic diversity that has been conserved in this species following the distributional contraction experienced at the beginning of the current interglacial period.

MATERIAL AND METHODS

Genetic data and analyses

Trollius europaeus samples were collected during summers 2006, 2007 and 2008 in 79 populations covering the whole European range of the plant (Figure 1, Appendix S1 in Supporting Information). Plant tissue from up to six individuals per population were conserved in silica gel and DNA was extracted using the standard protocol of the QIAGEN DNeasy 96 Plant kit (QIAGEN, Hombrechtikon, Switzerland). amplified fragment length polymorphism (AFLP; Vos *et al.*, 1995) was performed with *Eco*RI and *Mse*I endonucleases (Promega, Dübendorf, Switzerland) and further amplified with the two primer pairs E-ATC/M-CAT and E-ACA/M-CTG. For digestion, 5µl of DNA were added to a final volume of 20 µl, containing 1x Buffer 2, 0.1mg/ml BSA and 1 unit/µl of *Mse*I and *Eco*RI; the mix was incubated at 37°C for 2h. Adaptors were ligated to the digested products at 37°C for 2h, using 40µl of mix containing 1X Buffer T4, 0.45µM adaptor E, 0.36µM adaptor M, 0.015 units/µl T4 ligase (Promega, Dübendorf, Switzerland) and 20µl of digestion product. Preselective PCR was run in a 20µl mix composed of 1X buffer GoTaq (Promega SA, Dübendorf, Switzerland), 2µM $MgCl_2$ (Promega, Dübendorf, Switzerland), 250 µM dNTPs (Promega, Dübendorf, Switzerland), 0.25µM primers E-A and M-C, 0.025 units/µl of Go-Taq polymerase (Promega, Dübendorf, Switzerland) and 2µl ligated DNA. Thermocycler conditions were 2min at 94°C, followed by 29 cycles of 45 s at 94°C, 45 s at 56°C and 1 min

at 72°C, with 10 min of final extension at 72°C. The final selective step was done in 20µl mix composed of 1X buffer GoTaq (Promega, Dübendorf, Switzerland), 2µM $MgCl_2$ (Promega SA, Dübendorf, Switzerland), 250 µM dNTPs (Promega, Dübendorf, Switzerland), 0.4µM of selective primers E-ACA/M-CTG and E-ATC/M-CAT, 0.025 units/µl of Go-Taq polymerase and 3µl of a solution 1/20 of the preselective PCR product. Thermocycler conditions were 2 min at 94°C, followed by 13 cycles of 30 s at 94°C, 30 s at 65°C with temperature lowering 0.7°C at each cycle and 1 min at 72°C, followed by 23 cycles of 30 s at 94°C, 30 s at 65°C and 1 min at 72°C, and 5 min of final extension at 72°C. Selective amplification products were run by Macrogen Inc. (Seoul, South Korea) in a denaturing polyacrylamide gel with an internal size standard (ROX 400HD) on an automated DNA sequencer (ABI 377; Applied Biosystems, Foster City, USA). Automated allele-calling was further performed using Genemapper 3.7 (Applied Biosystems, Foster City, USA) with the following parameters: RFU threshold fixed to 50; binning range between 50bp and 270bp for E-ACA/M-CTG and 50bp and 350bp for E-ATC/M-CAT. Because digestion and amplification problems can arise during the procedure and can bias results (Arrigo *et al.*, 2009), we randomly chose 36 individuals per plate to be used as intra-plate replicates, while five samples out of the total were selected as inter-plate replicates. To test for allele reproducibility, we analyzed the percentage of bands shared by all replicates of a given sample.

In order to identify the genetic structure of the dataset, ten Monte-Carlo Markov chain (MCMC) runs of 1,000,000 generations with a 200,000 burnin period were run for an *a priori* number of genetic clusters (hereafter referred to as *K*) ranging from one to 20 (200 runs in total) in a model-based Bayesian framework implemented in STRUCTURE 2.2 (Falush *et al.*, 2007). Following recommendations of D. Falush (pers. comm.), data was coded as diploid and recessive alleles as present. The most probable *K* was identified using approaches proposed both by Pritchard *et al.* (2000) and Evanno *et al.* (2005).

In parallel and in order to confirm the genetic structure, a non-model-based approach was applied using the K-means clustering technique as introduced by Hartigan and Wong (1979) and as recently implemented in an AFLP-based phylogeographic framework by several

studies focusing on European plants (Burnier *et al.*, 2009; Arrigo *et al.*, 2010). Following Kergoat and Alvarez (2008), we identified the number of clusters (*K*) that optimizes the inertia (*i.e.*, the average distance between the centroid of each cluster and the samples assigned to each of them) of the dataset, and that best represents its most parsimonious splitting; the best *K* is identified by the analysis of the variation of inertias and of their first derivatives for each value of *K*. Here, inertias and assignments were calculated for *K* ranging from one to 20. Calculations were repeated starting at different random points 10,000 times, and run in the R 9.2.1 CRAN environment (R Core Development Group, 2009). Scripts are available upon request to the first author.

Nei genetic diversities (Nei & Li, 1979) were calculated based on samples having successfully amplified fragments using both AFLP primer-pairs, by applying a geographic moving window (Arrigo *et al.*, 2010) in a R 9.2.1 CRAN environment (R Core Development Group, 2009). The method consists of the following steps: i) the region of study is defined with maximum longitude and latitude values; ii) a grid of fixed cell-size (in our case 75 km) and their associated centroids are calculated; iii) a fixed maximum number of samples (four in our analysis) is randomly selected in a user-defined centroid perimeter (200 km in our case); iv) Nei's diversity indices are calculated for each centroid, v) points iii and iv are repeated 1,000 times, vi) mean values of the 1,000 Nei's indices are calculated for each centroid. This method allows not only to calculate the Nei diversity values at each point, but also to observe the index variation and distribution throughout the area of study. Moreover, it allows for a correction of possible local sampling biases since the analysis is grid-based, instead of sampled-population based. Scripts are available upon request to the first author.

Finally, to confirm the genetic structure obtained, a distance-based phylogenetic analysis was applied based on the calculation of modified Nei-Li pairwise distances between genotyped individuals, as implemented in PHYLIP ("restdist" and "neighbor" programs; Felsenstein, 2005). The resulting topology was rooted using the mid-point rooting method, as implemented in FigTree 1.2.2 (http://tree.bio.ed.ac.uk/software/figtree).

Environmental analyses

Ecological Niche Modelling (ENM) and hindcasting

Occurrences of *T. europaeus* recorded with a spatial accuracy ≤ 5km were extracted from seven biodiversity databases (GBiF: www.gbif.org; Flora Croatica: hirc.botanic.hr/fcd/Search.aspx; Swedish Artportalen: http://www.artportalen.se; database of the Botanical Alpine Conservatory of Gap-Charance, France; Anthos Project: www.anthos.es; NBN gateway: www.searchnbn.net/index_homepage/index.jsp; Centre du Réseau Suisse de Floristique www.crsf.ch) with a total of 49,863 occurrences. The accuracy threshold has been applied taking into account that 5km is also the approximate resolution of the most detailed climatic grids currently available (see below). This dataset was pooled with occurrences collected during fieldwork (79 locations throughout the distribution range, see above). Because the total 49,932 occurrences were highly aggregated in some regions of Europe (mostly western Europe), we avoided oversampling by defining a minimal distance between points (*i.e.*, 20 km) and randomly selected a subset of occurrences spatially distributed across the range of the species using this buffer.

Most modeling techniques require not only information on presences but also on absences or pseudo-absences when no true absences are available (Wisz & Guisan, 2009). Therefore, 10,000 pseudo-absence points were selected based on a random selection of pixels with a resolution equal to that of the climatic layers (2.5 arc-minutes). Because the databases used in the present study provided very good information on occurrences of the plant in the western and central parts of the sampling area, but not in most eastern parts, pseudo-absences were only defined within the limits of the investigated range (Figure 1) and not within the whole distribution of the plant, extending towards the Russian Urals. This was done to avoid biases in the response curves of the models (Phillips *et al.*, 2009).

Because biases in the ENM definition can arise when too many correlated climatic variables are considered (Nogués-Bravo, 2009), we selected the following seven WorldClim bioclimatic layers (Hijmans *et al.*, 2005), which are recognized to be of strong importance for the ecology of alpine and arctic plants (Guisan & Zimmermann, 2000; Körner, 2003):

minimum temperature (bio5); maximum temperature (bio6); average temperature of the coldest quarter (bio10); average temperature of the warmest quarter (bio11); precipitation seasonality (bio15); average precipitation of the coldest quarter (bio18); and average precipitation of the warmest quarter (bio19). The distribution of *T. europaeus* was modeled using eight of the modeling techniques proposed by the BIOMOD tool (Thuiller *et al.*, 2009), implemented in the R 9.2.1 CRAN environment (R Core Development Group, 2009): (1) generalized linear model (GLM), (2) generalized additive model (GAM), (3) classification tree analysis (CTA), (4) artificial neural networks (ANN), (5) multivariate adaptive regression splines (MARS), (6) mixture discriminant analysis (MDA) (7) generalized boosting model (GBM), and (8) random forest (RF).

In order to evaluate the predictive performance of the species distribution model, we used a random subset of 70% of the data to calibrate each model, while the remaining 30% was used for its assessment (following, Thuiller *et al.*, 2009). Models were evaluated using a Relative Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) (Fielding & Bell, 1997). We replicated 50 times the data splitting and calculated the average AUC of the repeated split-sample, which gives a more robust estimate of the predictive performance of each model (Thuiller *et al.*, 2009).

Finally, each model was projected into the past with the WorldClim data of the Community Climate System Model (CCSM) circulation model for the last glacial maximum (LGM; -21,000 years) downscaled at a resolution of 2.5 arc-minute. Because ensemble forecasting approaches significantly improve the accuracy of species distribution models (Marmion *et al.*, 2009), we applied a weighted average of the eight modeling techniques based on their predictive power (AUC). We thus obtained the main tendency of these distributions, accounting also for possible variation among modeling techniques.

Circumscription of regions of high hindcasted probabilities

ENM predicts habitat suitability (in our case based on climatic factors) of the study area for the modeled organism by attributing a probability value to each pixel. Since we aim at

delimitating regions of high suitability, it is necessary to define a threshold to retain only those regions in which the probability of suitability for long-term persistence is high. In order to do so, several methods have been previously proposed (Liu *et al.*, 2005; Jiménez-Valverde & Lobos, 2007). Some of these are based on solid statistical backgrounds and are useful when applied to pure theoretical modeling approaches, while others present a deeper biological relevance, more compatible with an evolutionary framework. This is why in the present study we applied two different approaches. First, the one proposed by Manel et al (2001), which considers both presences and absences and defines a threshold optimizing sensitivity/specificity ratios (referred to as "SS"). Second, taking into account only the variation in presence data, we extracted the current probability values predicted by the model for these points, and calculated their mean and standard error (SE). In this method (referred to as "M"), the threshold value corresponded to "mean - SE". With both "SS" and "M" methods, regions falling below thresholds were considered as marginal habitats.

Once cut-offs were applied, areas with high hindcasted probabilities, hereafter referred to as *suitability hotspots*, were defined according to the following rules: i) their surface is large and continuous (*i.e.*, we did not consider punctual regions of high probability if they occupied areas smaller than 1,000 km^2); ii) their distribution range is isolated from others hotspots (*i.e.*, by boundaries composed of large regions of low-hindcasted probabilities and/or by the presence of ice-sheets).

Inference of the post-glacial history

In order to identify an eventual post-glacial loss of diversity, we applied a strategy similar to those proposed by Magri *et al.* (2007) and Cheddadi *et al.* (2006), in terms of a complementary use of non-molecular data. In our case, the inference method replaced inferences on fossil data by an ENM approach, which allows extending the application range of the technique to virtually any cold-adapted species. We compared the number and distributions of current genetic clusters with those of *suitability hotspots* defined from the ENM approach, using the following inference key: i) if their numbers and respective

distributions were similar, the probability is high that at least one genetic lineage per *suitability hotspot* was preserved after the last postglacial contraction; ii) if the number of *suitability hotspots* was higher than the number of genetic groups, this would indicate, assuming that lineages were not shared between hotspots, a probable genetic loss due to postglacial contraction; iii) if the number of genetic groups was higher than the number of *suitability hotspots*, this would indicate that lineages from each ENM *suitability hotspots* participated to the current diversity and that at least one of them harbored more than one single lineage.

RESULTS

Genetic analyses

Among the 354 extracted samples of *T. europaeus*, 270 samples were successfully genotyped with the two primer pairs (374 AFLP fragments in total). In addition, 15 and 64 samples amplified correctly only for a single primer pair. Mismatch error rate (*sensu* Bonin *et al.* 2004) was 0.90% for primer-pair E-ACA/M-CTG and 0.91% for primer-pair E-ATC/M-CAT.

The spatial genetic structure inferred using STRUCTURE showed that for values of $K \geq 4$, the number of "realized" clusters (*i.e.*, comprising maximum probabilities higher than 0.005) was always equal to four and the assignment of samples to each group was identical in all cases. According to Pritchard *et al.* (2000), there is thus a strong structuring of the data into four clusters. Following Evanno *et al.* (2005), the most likely *K* value was nine, whose best run yielded four "realized" clusters and five "ghost" clusters (*i.e.*, comprising maximum probabilities lower than 0.005). The assignment of samples to the four clusters was strongly structured geographically (Figure 1a): the first cluster was exclusive to south-eastern Europe (blue); the second was only present in northern Scandinavia (green); the third was mainly restricted to south-western Europe (red); and the fourth was more widespread with a central/northwestern European distribution (yellow). At the edges of each cluster distribution, some admixture was observed.

343 The non-hierarchical, non model-based, K-means analysis also indicated that the *K* value
344 that optimized the distribution of inertias was four (Appendix S2 in Supporting Information).
345 When the assignment of samples to each genetic group was analyzed, results were almost
346 identical to those obtained in the model-based STRUCTURE analysis: significant positive
347 correlations between both methods for the assignment of samples to different clusters were
348 illustrated by R^2 values and slopes ranging from 0.84 to 0.96, and from 0.71 to 0.90, for each
349 *K* cluster (Appendix S3 in Supporting Information).
350 The geographic structure of the Nei diversity indices showed a main centre of diversity in
351 the southernmost regions (Alps), with additional highly diverse areas observed in northern
352 Carpathians, central Scandinavia and the North of the British Islands (Figure 1b). In contrast,
353 there seemed to be a loss of genetic diversity in northern continental Europe and southern
354 Scandinavia.
355 The phylogeny inferred from pairwise AFLP-based Nei-Li distances confirmed the
356 existence of the four genetic groups previously identified by the clustering approaches
357 (Figure 2), although their monophyly was not strictly retrieved. While south-western samples
358 were monophyletic and segregated basally, samples from central/north-western and south-
359 eastern Europe were paraphyletic, the later including a nested cluster comprising
360 monophyletic northern Scandinavian samples.
361
362 **ENM and hindcasting**
363 AUC values ranged between 0.772 and 0.847 (Table 1) indicating that the models were
364 performing well (i.e., >0.75; Swets, 1988). Current and past climatic suitability showed
365 substantial differences in their geographic extents (Figure 1c-d). The projection of the model
366 for the current environment (Figure 1d) indicated high probability of occurrence in southern
367 and central European mountain ranges (Pyrenees, Apennines, Balkans, Carpathians, Alps),
368 as well as at lower altitudes in northern Europe (Scandinavia, North of the British Islands). It
369 also predicted high probabilities in the north-eastern European wetlands (Baltic region) and
370 extended further East. The suitable area highlighted by the model fitted well with the actual

371 distribution of the plant (Meusel *et al.*, 1965). Hindcasted probabilities (Figure 1c) indicated
372 that a large extent of the land exposed during the LGM showed high suitability for *T.*
373 *europaeus*.

374 The two cut-off methods ("SS" and "M") applied provided similar results, even though "M"
375 appeared as more restrictive than "SS", with threshold values of 0.578 and 0.494,
376 respectively. After the application of the cut-off and identification of centres of high
377 probabilities at LGM, five *suitability hotspots* were observed (Figure 1c): i) western Pyrenees;
378 ii) north-eastern Europe, iii) western Europe, with a particularly high probability area in the
379 West (method "SS"); iv) central Carpathian lowlands; v) Northern Balkans with a probable
380 connection with the Apennines. In contrast, regions showing low occurrence probabilities
381 were mainly present in the northern Carpathians and surrounding areas, in central France
382 and across the southern tips of the Mediterranean peninsulas.

383

384 **Congruence between LGM *suitability hotspots* and current spatial genetic structure**

385 The combined evaluation of hindcasted probabilistic hotspots and current spatial genetic
386 structure showed that the number and localization of current genetic groups roughly
387 corresponded to regions of high hindcasted probabilities. We observed however, that there
388 was one additional hindcasted hotspot (five), compared to the number of genetic clusters
389 (four): the genetic cluster covering southeastern Europe (blue, Figure 1a) covered the areas
390 of the two hotspots *iv* and *v* (Figure 1c). In contrast, each of the three other clusters
391 corresponded to one single hotspot: the northern Scandinavian genetic group (green, Figure
392 1a) occupied a region covered with ice-sheets during the LGM, but placed in the
393 neighbourhood of the hindcasted hotspot *ii*, located in the direct southeastern vicinity of the
394 glacial ice-sheet (Figure 1c); the southwestern clade (red, Figure 1a) largely overlapped with
395 the location of hotspot *i* (Figure 1c); finally, the central/northwestern European clade (yellow,
396 in Figure 1a) showed a position similar to that of hotspot *iii* (Figure 1c).

397 At least one of the two hindcasted *suitable hotspots iv* or *v* (Figure 1c) seemed to have not
398 participated to the composition of the current genetic diversity of the species in the sampled

399 area, unless these two neighbouring areas shared identical lineages during LGM (see
400 Discussion).
401
402 **DISCUSSION**
403 **Spatial genetic structure**
404 The spatial genetic structure of *T. europaeus* identified in our study gives a much more
405 complete insight into its European phylogeographic pattern when compared to a previous
406 study on this plant (Després *et al.*, 2002), which comprised a much more scattered sampling
407 (18 sampling locations *vs.* 79 in the present survey). The major novelties and corrections
408 addressed here are (i) the presence of a cluster specific to southeastern Europe (blue,
409 Figure 1a), (ii) the admixed genetic identity of several locations in the southern Alps and
410 eastern Pyrenees, and (iii) the existence of an exclusive and independent northern
411 Scandinavian cluster (green, Figure 1a), different from the cluster found in southern
412 Scandinavia, the Carpathians and northern Poland (yellow, Figure 1a). Our results represent
413 the first evidence in arctic-alpine plants that Scandinavia was colonized twice by two distinct
414 lineages that established a contact zone at mid latitudes similar to the phylogeographic
415 pattern demonstrated for the bear *Ursus arctos* in this region (Hewitt, 1999), but contrasting
416 with results obtained in other cold-adapted species (in contrast to Schönswetter *et al.*, 2006;
417 Skrede *et al.*, 2006; Schmitt, 2009).
418 Overall, the spatial genetic structure inferred in this study shows a well circumscribed
419 geographic organization of the genetic variation whatever approach is considered (Figures
420 1a and 2). The phylogenetic analysis (Figure 2) provides results highly congruent with those
421 obtained in the clustering analysis (Figure 1a). Although the Nei-Li distance-based analysis
422 lacks sensitivity, notably when inferring branch lengths (Felsenstein, 2004), the mid-point
423 rooted topology is compatible with an organization of the genetic diversity into four genetic
424 groups. Nevertheless, only samples from the south-western region (red) were retrieved as
425 monophyletic and basally segregated from the other samples. Monophyletic North
426 Scandinavian samples (green) are nested within samples from the Carpathians/Balkans

region (blue). Similarly, samples from this latter area are nested within samples from central Europe and southern Scandinavia (yellow). The paraphyletic and nested nature of these groups might, however, be due to methodological biases related to distance calculations based on AFLP data.

Extrapolated grid-based Nei genetic diversities (Figure 1b) indicate that regions of high diversity are mainly distributed in the southern part of the distribution, especially in the Alpine arc (Figure 1b). The high diversity retrieved in southern massifs might be explained by the potential of these massifs to play a role of long-term refugia during interglacial periods (*sensu* Stewart *et al.*, 2010). The high level of diversity found all over the Alpine arc might also result from the presence of ancient (*i.e.*, pre-LGM) glacial contact zones that could have been "preserved" today in a refugial state (*in situ* survival, *sensu* Brochmann *et al.*, 2003). Alternatively, it could also reflect a colonization of this central mountain range by several extra-Alpine contracting genetic groups after the retreat of the ice-sheets which covered the Alps during the LGM, and thus created a current postglacial contact zone (i.e., habitat tracking hypothesis, *sensu* Brochmann *et al.*, 2003). The latter scenario is favoured by the spatial genetic structure retrieved here (Figure 1a), which points out the existence of three lineages (red, yellow and blue; Figure 1a) in this mountain range. The other three regions that also harbour high levels of genetic diversity are the Sudetes, central Scandinavia and northern Scotland. While the first two regions correspond to contact zones between neighbouring genetic lineages (yellow and blue clusters for the Sudetes, green and yellow clusters for central Scandinavia; Figure 1a), the third region represents a more unexpected case of admixture, which could be due to long-distance dispersal events of individuals from three different clusters that reached the north of the British Isles independently [i.e., three different lineages (red, yellow and blue; Figure 1a) are admixed in one single population in the Grampians Mounts].

General considerations on ENM and hindcasting for *T. europaeus*

Because (i) the current ENM-predicted suitable area for *T. europaeus* fits well the known distribution of the species (Meusel *et al.*, 1965) (Figure 1d) and (ii) the hindcasted LGM suitable area predicted (Figure 1c) is congruent with past distributions hindcasted for boreal trees such as *Betula pubescens* or *Salix caprea* (Svenning *et al.*, 2008) that currently occupy habitats similar to that of *T. europaeus* (European Comission DG Environment, 2007), it is highly probable that the hindcasted model accurately reflects the past distribution of the species. Once thresholds are applied, our results indicate that, as expected, the distribution range of *T. europaeus* was more extended during the LGM than it is today and covered a large part of the lands free of ice (Figure 1c). Several patches of the current distribution, particularly in southern Europe, were thus probably interconnected during the LGM (see below).

It is worth noting that the two threshold methods, “M” and “SS”, yielded very similar results, the first being slightly more restrictive. With the latter, several hindcasted *suitability hotspots* could have been interconnected, such as those corresponding to western Europe (*iii*) and the Pyrenees (*i*) or to the Carpathians (*iv*) and the Balkans (*v*). Finally, one should take into consideration that because during glacial cycles lowered sea levels rendered possible several dry land connections, a continuous distribution might have occurred between the Italian peninsula and both the Balkan (Mussi, 1990) and the West-Pyrenean (Rigaud & Simek, 1990) *suitable hotspots*, which would also coincide with our molecular results (Figure 1a).

Relationships between genetic clusters and hindcasted *suitability hotspots*

Our novel approach combines inferences on the current intra-specific genetic structure with information on hindcasted distributions and provides a new framework to understand the dynamics of cold-adapted distributional contractions. Using *T. europaeus* as a case study, we identified four genetic clusters (Figure 1a) and five *suitability hotspots* (Figure 1d), which roughly match spatially.

481 The south-western cluster (red, Figure 1a) is spread in the Pyrenees as well as in
482 surrounding mountain ranges such as the Cantabrics and the French Massif Central, and
483 extends further East towards the western Alps and the Appenines, where it is admixed with
484 other neighbouring lineages. The eastward spread of the cluster into this latter massif could
485 attest of the existence of a corridor that, at this time, might have linked the Italian peninsula
486 with southern France (Rigaud & Simek, 1990). The dominance of this cluster in several
487 southern massifs, which can be considered as contemporaneous refugia for cold-adapted
488 species (Stewart *et al.*, 2010), shows that this lineage, despite being currently highly spatially
489 fragmented, is not yet subdivided into more restricted lineages. This result gives clues on the
490 timing of differentiation: the 18,000 years since the LGM were not sufficient to allow detecting
491 for genetic differentiation. Because the distribution of this cluster fits well with the position
492 and extent of the *suitability hotspot i* (Figure 1c), it is likely that the latter harboured the
493 ancestors of lineages currently found in the Pyrenees, Massif Central, Cantabria and to some
494 extent the Apennines.

495 The central/northwestern cluster (yellow, Figure 1a) covers the central Alpine region as
496 well as southern and central Scandinavia, central Europe and the British Isles. Some
497 admixture is observed on the edges of its distribution, represented mainly by contact zones
498 with other neighbouring genetic groups, as in the Sudetes mountain range, the Massif
499 Central and the Southern edges of the Alps. This cluster is the currently most widespread
500 across Europe and the most frequent in the Alpine arc. Explanations for this pattern could be
501 as follows: (i) this genotype was the most widely distributed across the European steppe
502 during the LGM; (ii) it was the most central and thus able to quickly colonize the newly
503 deglaciated lands (Scandinavia and the Alps) (Hewitt, 1996); (iii) a combination of (i) and (ii).
504 The distribution of this cluster fits well the *suitability hotspot iii* (Fig. 1c), located close to the
505 southern edge of the northern ice-sheet, which could represent the LGM origin of this
506 lineage. Its position and extent could largely explain the presence of the yellow cluster in the
507 British Isles, as well as in Southern Scandinavia. Accordingly, when the ice-sheets started to

melt down, invasion of new free lands in northern Europe would have happened more easily from this hotspot than from any other.

The blue cluster (Figure 1a) is restricted to the Carpathians, the Sudetes, the Balkans and to a lesser extent the south-eastern Alps and the Apennines. The presence of this cluster in the latter massif (although admixed with the red cluster) is probably a consequence of land exposure of part of the Adriatic sea due to lowered sea levels during the LGM, which allowed the communication between the Italian peninsula and the Balkans (Mussi, 1990). This cluster does not extend to the North further than the Sudetes, as shown in other Eurasian organisms, in which the Carpathians harbour specific lineages, as in the cases of the plant *Arabis alpina* (Ehrich *et al.*, 2007) or the butterfly *Erebia euryale* (Schmitt & Haubrich, 2008). This pattern is expected since the northern Carpathians show a longitudinal extension along a West/East axis, which is known to have provided shelter against the cold and dry winds coming from the northern plains during glacial periods (Feurdean, 2004), but which also rendered dispersal towards the North more difficult. Accordingly, current Poland and western Ukraine were characterized by LGM climatic conditions too extreme to allow the survival of *T. europaeus*, a species adapted to humid habitats. As indicated by the hindcasted ENM, the Carpathian and Balkan regions appear to have provided suitable climatic conditions for *T. europaeus* (Figure 1c), harbouring two *suitability hotspots* (*iv* and *v*) strongly isolated towards the North. However, because only one genetic cluster was found in this region, we cannot identify which of these two regions was the most probable source for the blue cluster. It is also possible that the absence of a second genetic cluster in this area is a consequence of the geographical proximity of these two hotspots, which might have experienced continuous gene flow during glacial periods. Such a homogenising force would thus have limited *in situ* diversification in each hotspot.

Finally, the location of the North-Scandinavian cluster (green, Figure 1a) suggests that this lineage invaded the region from a zone outside the sampled range. Good candidates of origin for this group are the Russian plains, an area known to have acted as a colonization source of the whole Scandinavian peninsula in other organisms (e.g. Brochmann *et al.*, 2003;

Alsos *et al.*, 2009). Including plant samples from the Russian edge of the distribution range of the plant would allow testing this hypothesis. The hypothesis of an eastern colonization of northern Scandinavia is also compatible with results obtained by the hindcasting approach, which identified the *suitability hotspot ii* in the south-eastern edge of the northern ice-sheet (Figure 1d) that could have extended into the Russian plains. This hotspot area thus appears as the most likely origin of the green cluster, which would have invaded northern Scandinavia once the northern glacial sheets retracted.

Preservation or loss of genetic diversity?

In contrast to temperate species, cold-adapted species show refugial contracted distributions during interglacial periods (Schmitt, 2009). In Europe, long-term and cryptic refugia have been identified for cold-adapted arctic-alpine organisms, for instance in Scandinavia and in the southern mountain ranges, respectively (Stewart *et al.*, 2010). The distribution of *T. europaeus* currently covers most areas harbouring one or the other type of these refugia.

Based on the combination of the two independent analytical approaches presented here, we can propose a post-glacial contraction scenario for *T. europaeus* (Figure 3), in which four contraction/colonization events, mostly in a centrifugal manner (*i.e.*, from central lowlands to surrounding highlands), explain the current spatial genetic structure found in the species.

As shown in Figures 1 and 3, the number of genetic clusters is lower than that of *suitability hotspots* during the LGM. Based on the parsimony criterion it seems plausible that all but one *suitability hotspots* participated to the current genetic diversity since in the area comprising the Balkans (*suitability hotspot v*) and the Carpathians (*suitability hotspot iv*) only one genetic cluster was retrieved (see grey crossed zone in Figure 3). Past genotypes associated with one of these two hindcasted hotspots could have been lost during post-glacial contraction or are currently restricted to highly localized areas, not covered by the present study (*e.g.*, current Montenegro and Bosnia-Herzegovina). It is also likely that the distance between the Carpathian and Balkan *suitable hotspots* was small enough to allow gene flow and that these

two regions finally shared the same lineage. As a consequence, the loss of diversity in *T. europaeus* during the interglacial contraction period seems to be reasonably low, especially when compared to results obtained for organisms such as the arctic fox, which lost a substantial proportion of its Pleistocene genetic diversity after the end of the LGM (i.e., all European and some Scandinavian lineages, Dalén *et al.*, 2007).

In the present work, we demonstrate that the inference of postglacial dynamics of cold-adapted species, for which no ancient DNA is available, is possible by the simultaneous use of hindcasting and current spatial genetic structure analyses. In the future, and in order to test more accurately for the post-LGM contraction addressed in this study, additional coalescent-based approaches should be performed and the likelihood of different scenarios should be tested. This, however, would require an accurate evaluation of parameters related to population demography, a challenge that has not been overcome yet when using AFLP data (e.g., Beerli & Felsenstein, 1999; Excoffier *et al.*, 2005).

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Triponez, Dr. H. Hipperson, T. Suchan, M. Ronikier, N. Magrou, R. Lavigne, L. Juillerat, R. Arnoux, N. Villard, A. Sarr, P. Lazarevic, N. Russell, P. Bowler, R. Dafydd, Dr. M. Howe and D. Gyurova for their invaluable field and technical assistance, as well as member of the Carstens' laboratory whose comments largely improved an earlier version of this manuscript. This project was funded by the Swiss National Science Foundation (project No. 3100A0-116778).

REFERENCES

Alsos, I.G., Alm, T., Normand, S. & Brochmann, C. (2009) Past and future range shifts and loss of diversity in dwarf willow (*salix herbacea* l.) inferred from genetics, fossils and modelling. *Global Ecology and Biogeography*, **18**, 223-239

Alvarez, N., Thiel-Egenter, C., Tribsch, A., Holderegger, R., Manel, S., Schonswetter, P., Taberlet, P., Brodbeck, S., Gaudeul, M., Gielly, L., Kupfer, P., Mansion, G., Negrini, R., Paun, O., Pellecchia, M., Rioux, D., Schupfer, F., Van Loo, M., Winkler, M.,

593 Gugerli, F. & Intrabiodivconsortium (2009) History or ecology? Substrate type as a
594 major driver of spatial genetic structure in alpine plants. *Ecology Letters*, **12**, 632-640
595 Arrigo, N., Felber, F., Parisod, C., Buerki, S., Alvarez, N., David, J. & Guadagnuolo, R.
596 (2010) Origin and expansion of the allotetraploid *aegilops geniculata*, a wild relative of
597 wheat. *New Phytologist*, **187**, 1170-1180
598 Arrigo, N., Tuszynski, J.W., Ehrich, D., Gerdes, T. & Alvarez, N. (2009) Evaluating the impact
599 of scoring parameters on the structure of intra-specific genetic variation using
600 rawgeno, an r package for automating aflp scoring. *BMC Bioinformatics*, **10**, 33
601 Avise, J.C. (2009) Phylogeography: Retrospect and prospect. *Journal of Biogeography*, **36**,
602 3-15
603 Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A. &
604 Saunders, N.C. (1987) Interspecific phylogeography: The mitochondrial DNA bridge
605 between population genetics and systematics. *Annual Review of Ecology and*
606 *Systematics*, **18**, 489-522
607 Beerli, P. & Felsenstein, J. (1999) Maximum-likelihood estimation of migration rates and
608 effective population numbers in two populations using a coalescent approach.
609 *Genetics*, **152**, 763-773
610 Biodiversity, E.C.D.E.N.A. (2007) Interpretation manual of european union habitats - eur27.
611 In:
612 Brochmann, C., Gabrielsen, T.M., Nordal, I., Landvik, J.Y. & Elven, R. (2003) Glacial survival
613 or *tabula rasa*? The history of north atlantic biota revisited. *Taxon*, **52**, 417-450
614 Burnier, J., Buerki, S., Arrigo, N., Kupfer, P. & Alvarez, N. (2009) Genetic structure and
615 evolution of alpine polyploid complexes: *Ranunculus kuepferi* (ranunculaceae) as a
616 case study. *Molecular Ecology*, **18**, 3730-3744
617 Cheddadi, R., Vendramin, G.G., Litt, T., François, L., Kageyama, M., Lorentz, S., Laurent, J.-
618 M., De Beaulieu, J.-L., Sadori, L., Jost, A. & Lunt, D. (2006) Imprints of glacial refugia
619 in the modern genetic diversity of *pinus sylvestris*. *Global Ecology and Biogeography*,
620 **15**, 271-282
621 Dalén, L., Nyström, V., Valdiosera, C., Germonpré, M., Sablin, M., Turner, E., Angerbjörn, A.,
622 Arsuaga, J.L. & Götherström, A. (2007) Ancient DNA reveals lack of postglacial
623 habitat tracking in the arctic fox. *Proceedings of the National Academy of Sciences of*
624 *the United States of America*, **104**, 6726-6729
625 Després, L., Loriot, S. & Gaudeul, M. (2002) Geographic pattern of genetic variation in the
626 european globeflower *trollius europaeus* l. (ranunculaceae) inferred from amplified
627 fragment length polymorphism markers. *Molecular Ecology*, **11**, 2337-2347
628 Ehlers, J. & Gibbard, P.L. (2007) The extent and chronology of cenozoic global glaciation.
629 *Quaternary International*, **164-165**, 6-20

630 Ehrich, D., Gaudeul, M., Assefa, A., Koch, M.A., Mummenhoff, K., Nemomissa, S. &
631 Brochmann, C. (2007) Genetic consequences of pleistocene range shifts: Contrast
632 between the arctic, the alps and the east african mountains. *Molecular Ecology*, **16**,
633 2542-2559
634 Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals
635 using the software structure: A simulation study. *Molecular Ecology*, **14**, 2611-2620
636 Excoffier, L., Laval, G. & Schneider, S. (2005) Arlequin ver. 3.0: An integrated software
637 package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**,
638 47-50
639 Falush, D., Stephens, M. & Pritchard, J. (2007) Inference of population structure using
640 multilocus genotype data: Dominant markers and null alleles. *Molecular Ecology*
641 *Notes*, **7**, 895-908
642 Felsenstein, J. (2004) *Inferring phylogenies*. Sinauer Associates, Sunderland, MA.
643 Felsenstein, J. (2005) Phylip (phylogeny inference package) version 3.6. In. Distributed by
644 the author, Seattle
645 Feurdean, A. (2004) Palaeoenvironment in north-western romania during the last 15,000
646 years. In: *Department of Physical Geography and Quaternary Geology*. Stockholm
647 University, Stockholm
648 Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors
649 in conservation presence/absence models. *Environmental Conservation*, **24**, 38-49
650 Frenzel, B. (1968) Pleistocene vegetation of northern eurasia - recent vegetation of northern
651 eurasia resulted from a relentless contest between steppe and forest. *Science*, **161**,
652 637
653 Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology.
654 *Ecological Modelling*, **135**, 147-156
655 Hampe, A. & Petit, R.J. (2005) Conserving biodiversity under climate change: The rear edge
656 matters. *Ecology Letters*, **8**, 461-467
657 Hartigan, J.A. & Wong, M.A. (1979) A k-means clustering algorithm. *Applied Statistics*, **28**,
658 100-108
659 Hewitt, G.M. (1996) Some genetic consequences of ice ages, and their role in divergence
660 and speciation. *Biological Journal of the Linnean Society*, **58**, 247-276
661 Hewitt, G.M. (1999) Post-glacial re-colonization of european biota. *Biological Journal of the*
662 *Linnean Society*, **68**, 87-112
663 Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high
664 resolution interpolated climate surfaces for global land areas. *International Journal of*
665 *Climatology*, **25**, 1965-1978

666 Hutchinson, G.E. (1957) Cold spring harbor symposia on quantitative biology. Concluding
667 remarks. In: *Cold Spring Harbor Symposia on Quantitative Biology*, pp. 415-427
668 Jiménez-Valverde, A. & Lobos, J.M. (2007) Threshold criteria for conversion of probability of
669 species presence to either–or presence–absence. *Acta Oecologica-International*
670 *Journal of Ecology*, **31**, 361-369
671 Kergoat, G. & Alvarez, N. (2008) Assessing the phylogenetic usefulness of a previously
672 neglected morphological structure through elliptic fourier analyses: A case study in
673 *bruchus* seed-beetles (coleoptera: Chrysomelidae: Bruchinae). *Systematic*
674 *Entomology*, **33**, 289-300
675 Körner, C. (2003) *Alpine plant life*, 2nd edition edn. Springer, Heidelberg.
676 Liu, C., Berry, P.M., Dawson, T.P. & Pearson, R.G. (2005) Selecting thresholds of
677 occurrence in the prediction of species distributions. *Ecography*, **28**, 385-393
678 Magri, D., Fineschi, S., Bellarosa, R., Buonamici, A., Sebastiani, F., Schirone, B., Simeone,
679 M.C. & Vendramin, G.G. (2007) The distribution of quercus suber chloroplast
680 haplotypes matches the palaeogeographical history of the western mediterranean.
681 *Molecular Ecology*, **16**, 5259-5266
682 Manel, S., Williams, H.C. & Ormerod, S.J. (2001) Evaluating presence-absence models in
683 ecology: The need to account for prevalence. *Journal of Applied Ecology*, **38**, 921-
684 931
685 Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K. & Thuiller, W. (2009) Evaluation of
686 consensus methods in predictive species distribution modeling. *Diversity and*
687 *Distributions*, **15**, 59-69
688 Médail, F. & Diadema, K. (2009) Glacial refugia influence plant diversity patterns in the
689 mediterranean basin. *Journal of Biogeography*, **36**, 1333-1345
690 Meusel, H., Jäger, E. & Weinert, E. (1965) *Vergleichende chorologie der*
691 *zentraleuropäischen flora. Bd.1 (text und karten)*. Fischer, Jena.
692 Mussi, M. (1990) Continuity and change in italy at the last glacial maximum. *The world at*
693 *18000 bp* (ed. by O. Soffer and C. Gamble), pp. 126-147. Unwin Hyman, London.
694 Nei, M. & Li, W.H. (1979) Mathematical model for studying genetic variation in terms of
695 restriction endonucleases. *Proceedings of the National Academy of Sciences of the*
696 *United States of America*, **76**, 5269-5273
697 Nogués-Bravo, D. (2009) Predicting the past distribution of species climatic niches. *Global*
698 *Ecology and Biogeography*, **18**, 521-531
699 Phillips, S.J., Dudik, M., Elith, J., Graham, C., Lehmann, A., Leathwick, J. & Ferrier, S. (2009)
700 Sample selection bias and presence-only models of species distributions:
701 Implications for selection of background and pseudo-absences. *Ecological*
702 *Applications*, **19**, 181-197

703 Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using
704 multilocus genotype data. *Genetics*, **155**, 945-959
705 Pulliam, H.R. (2000) On the relationship between niche and distribution. *Ecology Letters*, **3**,
706 349-361
707 Rigaud, J.-P. & Simek, J.F. (1990) The last pleniglacial in the south of france (24000-14000
708 years ago). *The world at 18000 bp* (ed. by O. Soffer and C. Gamble), pp. 69-86.
709 Unwin Hyman Ltd, London.
710 Schmitt, T. (2007) Molecular biogeography of europe: Pleistocene cycles and postglacial
711 trends. *Frontiers in Zoology*, **4**, 11
712 Schmitt, T. (2009) Biogeographical and evolutionary importance of the european high
713 mountain systems. *Frontiers in Zoology*, **6**, 9
714 Schmitt, T. & Haubrich, K. (2008) The genetic structure of the mountain forest butterfly *erebia*
715 *euryale* unravels the late pleistocene and postglacial history of the mountain forest
716 biome in europe. *Molecular Ecology*, **17**, 2194-2207
717 Schönswetter, P., Popp, M. & Brochmann, C. (2006) Rare arctic-alpine plants of the
718 european alps have different immigration histories: The snow bed species *minuartia*
719 *biflora* and *ranunculus pygmaeus*. *Molecular Ecology*, **15**, 709-720
720 Skrede, I., Eidesen, P.B., Portela, R.P. & Brochmann, C. (2006) Refugia, differentiation and
721 postglacial migration in arctic-alpine eurasia, exemplified by the mountain avens
722 (*dryas octopetala* l.). *Molecular Ecology*, **15**, 1827-1840
723 Stewart, J.R., Lister, A.M., Barnes, I. & Dalén, L. (2010) Refugia revisited: Individualistic
724 responses of species in space and time. *Proceedings of the Royal Society of London*
725 *B*, **277**, 661-671
726 Svenning, J.-C., Normand, S. & Kageyama, M. (2008) Glacial refugia of temperate trees in
727 europe: Insights from species distribution modelling. *Journal of Ecology*, **96**, 1117-
728 1127
729 Swets, J.A. (1988) Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285-1293
730 Taberlet, P., Fumagalli, L., Wust-Saucy, A.G. & Cosson, J.F. (1998) Comparative
731 phylogeography and postglacial colonization routes in europe. *Molecular Ecology*, **7**,
732 453-464
733 Thuiller, W., Lafourcade, B., Engler, R. & Araújo, M.B. (2009) Biomod-a platform for
734 ensemble forecasting of species distributions. *Ecography*, **32**, 369-373
735 Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van De Lee, T., Hornes, M., Frijters, A., Pot,
736 J., Peleman, J. & Kuiper, M. (1995) Aflp: A new technique for DNA fingerprinting.
737 *Nucleic Acids Research*, **23**, 4407-4414

738 Wisz, M.S. & Guisan, A. (2009) Do pseudo-absence selection strategies influence species
739 distribution models and their predictions? An information-theoretic approach based on
740 simulated data. *BMC Ecology*, **9**, 8

741

742 **BIOSKETCH**

743 Anahí Espíndola is a PhD student at the University of Neuchâtel (Switzerland). She is
744 interested on the evolutionary history of interspecies interactions. Her research integrates
745 spatial, historical and ecological approaches to answer general phylogeographic and
746 biological questions. Loïc Pellissier is a PhD student at the University of Lausanne
747 (Switzerland) and currently works on ecological niche modelling approaches applied on
748 conservation of Alpine butterfly communities. Nadir Alvarez is a research associate at the
749 University of Lausanne, interested in the ecology and evolution of biological interactions.

750

751 Table 1 - Area under the curve (AUC) values for the different BIOMOD modeling techniques.

752

Modeling technique	AUC
Artificial Neural Networks (ANN)	0.797
Classification Tree Analysis (CTA)	0.815
Generalized Additive Model (GAM)	0.829
Generalized Boosting Model (GBM)	0.841
Generalized Linear Model (GLM)	0.826
Multivariate Adaptive Regression Splines (MARS)	0.826
Mixture Discriminant Analysis (MDA)	0.772
Random Forest (RF)	0.847

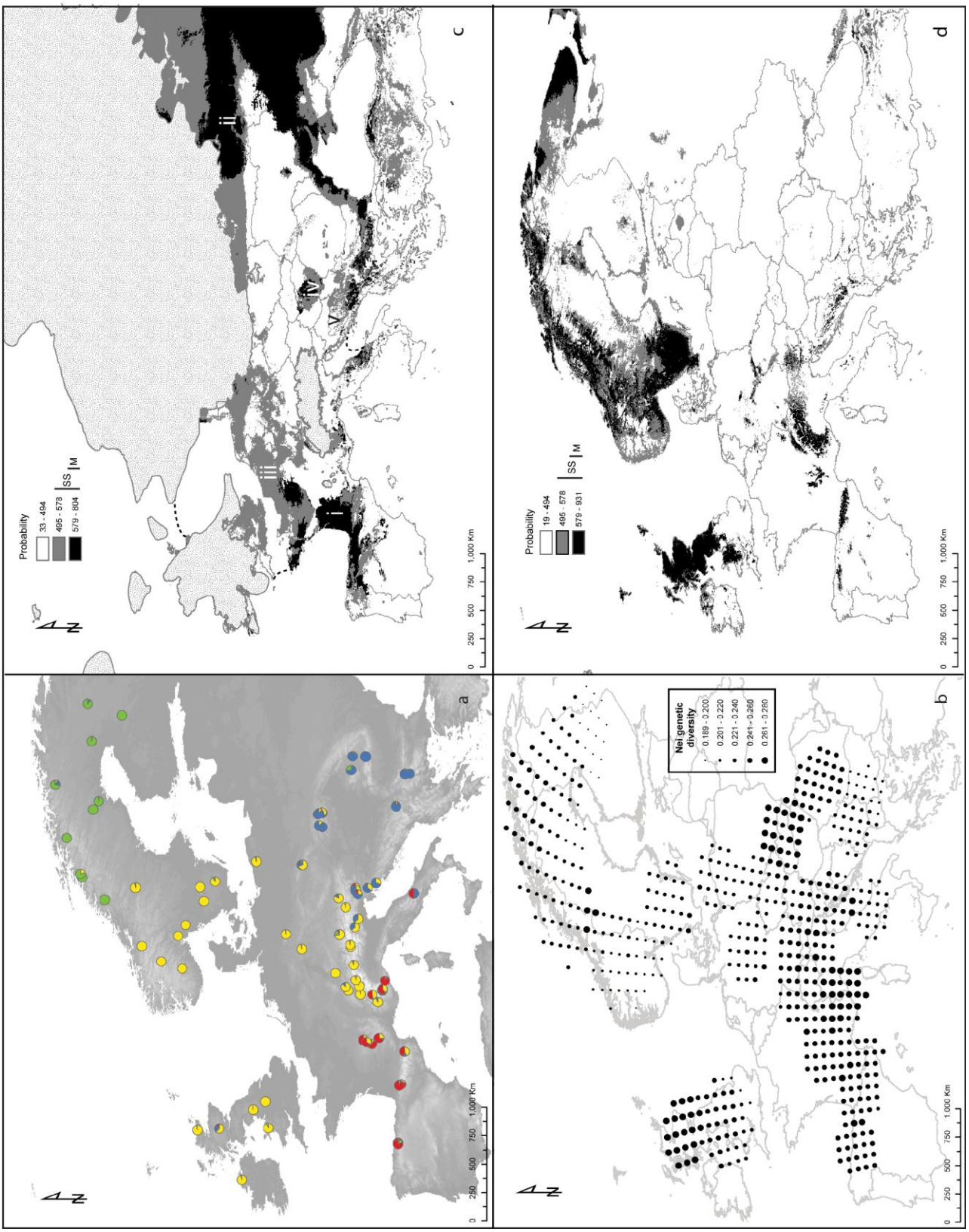
753
754

Figure captions

Figure 1 – Genetic structure (a) and estimated Nei genetic diversities (b) of *Trollius europaeus* in Europe. Projected climatic probabilities for past (c) and current (d) conditions, as inferred from the Ecological Niche Model (ENM) and after the application of the two thresholds ("M" and "SS"; see text). a – proportions of the pies indicate the probability to belong to one of four genetic clusters (colours), as estimated by STRUCTURE 2.2 (Pritchard *et al.*, 2000). b –dot sizes are proportional to the Nei genetic diversities extrapolated using a moving-window approach; scales are referred in the figure legend. c – d – probability values are indicated on each figure. White: values below both thresholds; grey: values below "M" threshold; black: values above both thresholds; in c, dotted areas correspond to glaciated regions; dashed lines define LGM coasts, as delimitated by Ehlers and Gibbard (2007); the five defined *suitable hotspots* are illustrated by roman numbers.

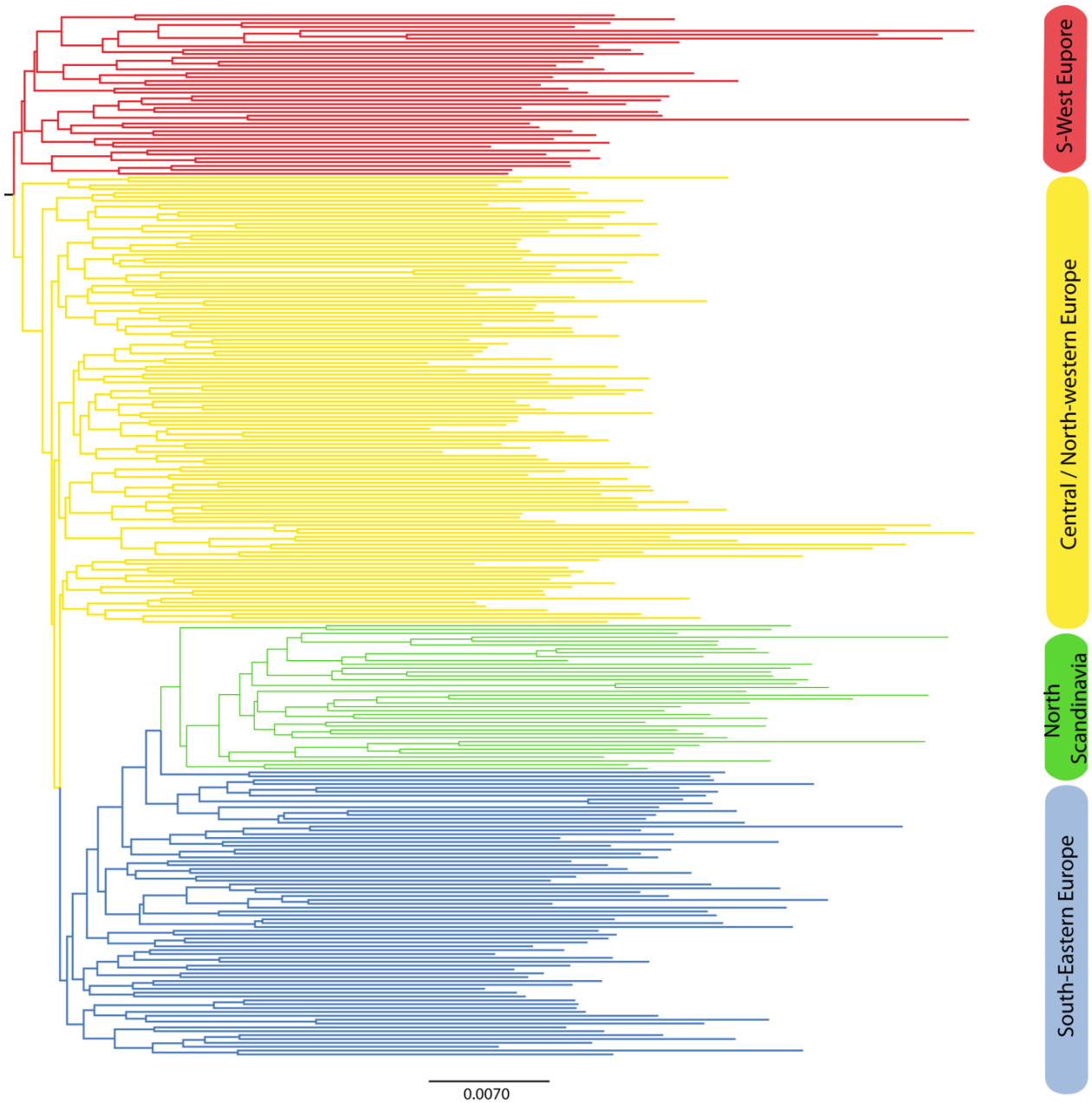
Figure 2 – Midpoint-rooting tree reconstructed with AFLP-based Nei-Li distances and Neighbour-Joining reconstruction for *Trollius europaeus*. Colours indicate regional clusters as retrieved in Figure 1a.

Figure 3 – Post-glacial contraction route scenarios for *Trollius europaeus*. Colours correspond to different genetic groups as in Figures 1 and 2; arrows indicate the direction of movement; thick dashed lines show main contact zones. The grey zone delimited by thin dashed lines indicates the putative region for which no genetic group was identified.


a
b
c
d
Nei genetic diversity
0.189 - 0.200
0.201 - 0.220
0.221 - 0.240
0.241 - 0.260
0.261 - 0.280
Probability
33 - 494
495 - 578
579 - 804
SS
M
Probability
19 - 494
495 - 578
579 - 931
SS
M
I
II
III
IV
V
0
250
500
750
1,000 Km

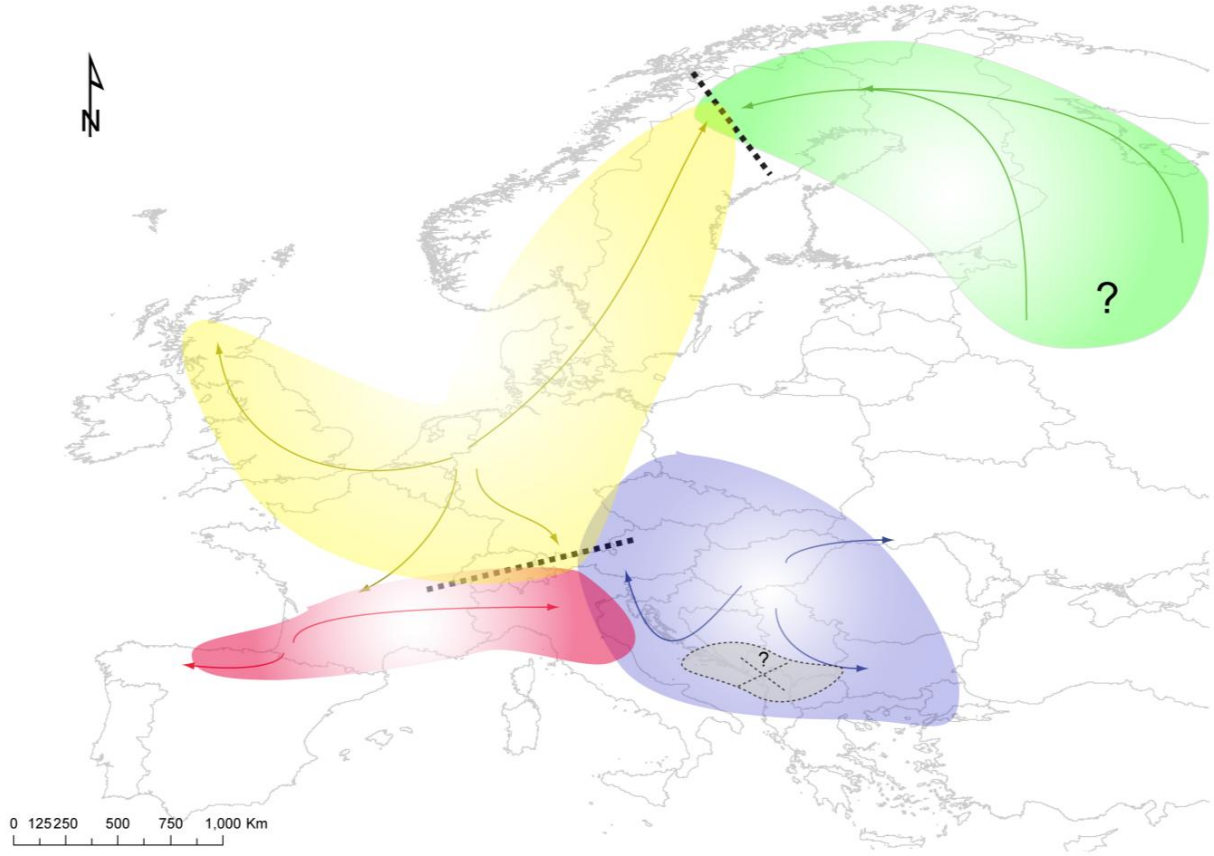
778

779


S-West Eupore
Central / North-western Europe
North Scandinavia
South-Eastern Europe
0.0070

780

781



782

783

784 Appendix S1 – Locations of sampled sites

Location	Country	Longitude	Latitude (m)	Altitude
Ambri	Switzerland	8.70292	46.50680	1000
Amot	Norway	8.42346	59.62199	481
Badgastein	Austria	13.09639	47.12056	1963
Bayasse	France	6.74067	44.30814	N/A
Beistohlen	Norway	8.95473	61.20761	729
Bidjovagge	Norway	22.47808	69.29778	613
Bobolice	Poland	16.59895	53.94660	110
Braas	Sweden	15.06817	57.09309	209
Chasseral	Switzerland	7.02130	47.12569	N/A
Chemin	Switzerland	7.08978	46.08993	N/A
Col de Bonnecombe	France	3.11410	44.57557	1335
Col de la Colombiere	France	6.46972	45.98722	1600
Col de Roanza	Italy	12.20989	46.17638	816
Col du Festre	France	5.86278	44.66639	1430
Col du Galibier	France	6.43861	45.08528	2000
Col long de Magnabaigt	France	-0.43646	42.87060	1615
Colt Park	England	-2.35247	54.19365	380
Crans-Montana	Switzerland	7.53890	46.34650	N/A
Cressbrook Dale	England	-1.74041	53.26724	246
Creux du Van	Switzerland	6.74119	46.93526	N/A
Donovaly	Slovakia	19.23068	48.88922	1345
Eidda Pastures	Wales	-3.74190	53.03720	234
Ellingsrudelva	Norway	10.91844	59.91771	161
Esposouille	France	2.09450	42.62341	1521
Fanske	Norway	15.40131	67.26452	46
Feleacu	Romania	23.59028	46.70118	670
Froson	Sweden	14.60268	63.18205	424
Glen Fender	Scotland	-3.79485	56.78138	345
Haiduta Mt	Bulgaria	23.32314	42.19968	2430
Haute Tinee 1	France	6.81871	44.29617	1770
Haute Tinee 2	France	6.85581	44.28426	1770
Hlupy	Slovakia	20.21572	49.24040	1775
Kajaani	Finland	27.77275	64.22848	141
Karersee	Italy	11.57341	46.40885	N/A
Karersee	Italy	11.57341	46.40885	N/A
Karlow	Poland	16.34070	50.47810	759
Krasno Polje	Croatia	14.97021	44.80746	1444
Lac Balea	Romania	24.62192	45.60239	2224
Laktatjakka	Sweden	18.40674	68.42931	453
Logarska Dolina	Slovenia	14.58630	46.36956	1401
Loser	Austria	13.78485	47.66052	1598

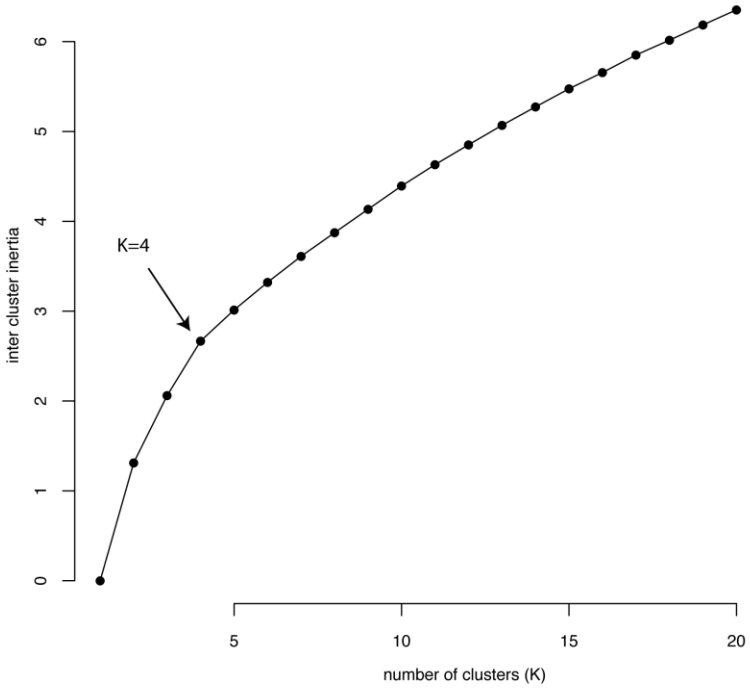
Location	Country	Longitude	Latitude (m)	Altitude
Lough Fern	North Ireland	-7.71130	55.06569	26
Moerlimatt	Germany	8.07760	47.90597	939
Monte Pizi	Italy	14.16714	41.91524	1546
Muntii Fagarasului	Romania	24.63611	46.59972	2205
Naverdal	Norway	10.13002	62.70417	480
O. Olteid	Norway	11.74560	59.34895	104
Pajino Preslo	Serbia	20.81970	43.27799	1802
petit papa noel	Finland	25.79386	66.51647	72
Pila	Slovakia	20.29449	48.90017	972
Podlesok	Slovakia	20.35190	48.94962	579
Puerto de Panderruedas	Spain	-4.97223	43.12743	1227
Puy de Dome	France	2.96333	45.77222	1460
Puy de Sancy	France	2.80972	45.53500	1520
Puy Mary	France	2.68083	45.11139	1550
Radkow	Poland	16.35321	50.46866	712
Risnjak - Snjeznik	Croatia	14.58494	45.43871	1466
Salla	Finland	28.65427	66.83020	194
Sede de Pan	France	-0.48651	43.03949	1556
Seterasen	Norway	13.67744	65.53432	285
Sjostorp	Sweden	14.65743	58.24696	100
Solberga	Norway	13.56116	57.95194	239
Sommerberg	Germany	9.92068	50.47913	770
Sta Marina de Valdeon	Spain	-4.87745	43.12790	1294
Steingaden	Germany	11.01296	47.59529	1158
Storzic	Slovenia	14.39488	46.34390	N/A
Strath Hallerdale	Scotland	-3.89403	58.45140	36
Straumen	Norway	15.64921	67.38440	75
Susch	Switzerland	10.07473	46.74728	N/A
Svartla	Sweden	21.22062	65.99583	36
Tarasp	Switzerland	10.25056	46.77730	N/A
Tatranska Lomnica	Slovakia	20.27944	49.16582	839
Tende	France	7.49970	44.10770	1622
Trollblumenwiese	Germany	11.04118	51.68314	488
Val de Choc	Slovakia	19.34333	49.15083	1550
Vitosha	Bulgaria	23.29342	42.59032	1779
Vuollerim	Sweden	20.57896	66.42037	166
Zali Log	Slovenia	14.11080	46.20342	533
Zielowe Ludowe	Serbia	16.34920	50.40210	695

785

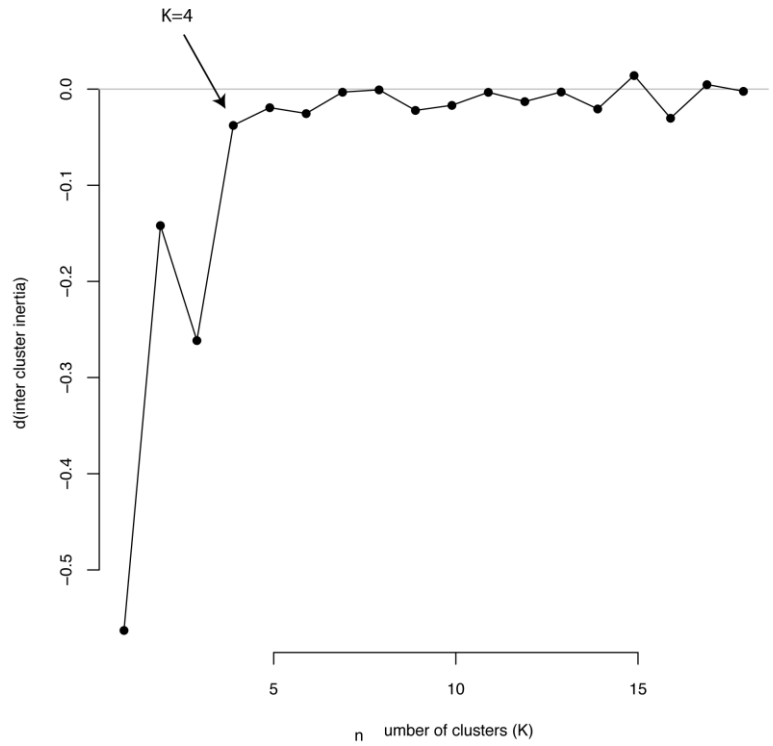
786

Appendix S2 – First (a) and second (b) derivatives for inertias, for all values of K calculated using the K-means approach.

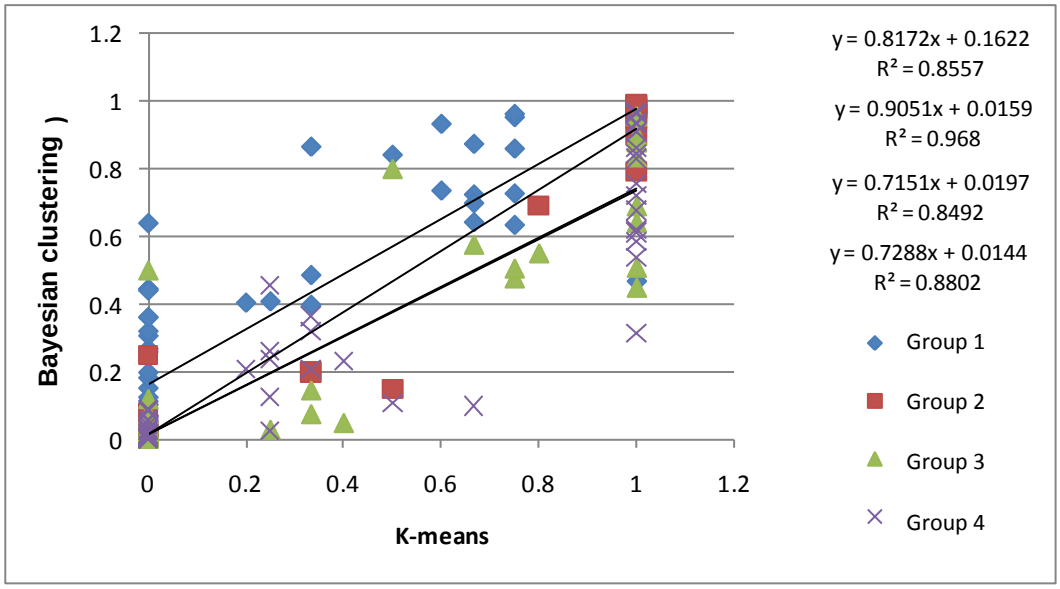
a)



b)



Appendix S3 –Correlation between sample assignments considering the two genetic clustering methods (Bayesian and K-means) used in this study.



Chapter eight

Phylogeography of mutualists and the effect of the host on the genetic structure of its partners

Anahí Espíndola, Bryan C. Carstens and Nadir Alvarez

Phylogeography of mutualists and the effect of the host on the genetic structure of its partners

Anahí Espíndola[1], Bryan C. Carstens[2] and Nadir Alvarez[3]

[1]Laboratory of Evolutionary Entomology, Institute of Biology, University of Neuchâtel. Emile-Argand 11, 2000 Neuchâtel, Switzerland.

[2]Life Sciences Building, Louisiana State University, Baton Rouge, LA 70803 USA.

[3]Department of Ecology and Evolution, University of Lausanne, Biophore Building, 1015 Lausanne, Switzerland

Abstract

The field of phylogeography has seen an important development during the last decades, especially in Europe and North America. In spite of this, little is known about the effect of ecological interactions on the phylogeographic patterns of interacting organisms. However, the coevolutionary theory predicts that interacting species should present similar histories, especially if interactions are specific and obligate.

To investigate this idea, we use here the mutualistic nursery pollination system formed by the Eurasian arctic-alpine species*Trollius europaeus* L. and its specific pollinators of the genus *Chiastocheta* Pokorny as a case study. In this system, the plant is visited and pollinated exclusively by the flies, whose larvae feed specifically on the plant seeds. Because the system concerns a low number of interacting species, it is highly appropriate to test the effect of interactions in a comparative framework. Based on a large-scale sampling of pollinating flies, we first infer the phylogeographic histories of each fly species from mitochondrial data. Because the large-scale genetic structure of the plant has been recently identified, we exploit this information and ask whether or not the phylogeographic patterns of the flies are significantly different of what would be expected under a scenario of plant-insect congruence.

Our results demonstrate that while all fly species occupy an extremely close ecological niche, their phylogeographic histories highly differ. While *C. rotundiventris* presents no geographic structure, *C. lophota* has a genetic structure largely explained by geography, while the phylogeographic pattern of *C. dentifera* does not statistically differ of what would be expected if it had evolved under a scenario guided by the plant. These results indicate that forces shaping the genetic structure of the three species are highly different. Differential dispersion capabilities could explain these patterns.

Introduction

The high number of species and enhanced rate of diversification in insects and angiosperms is often explained by reciprocal adaptive radiation of these two groups in a coevolutionary context (Simpson 1953; Schluter 2000; Lunau 2004; Futuyma and Agrawal 2009). In the coevolutionary model of Ehrlich & Raven (1964), the “escape and radiate” process enhances codiversification of plants and insects ecologically related by herbivory. Adjustments against exploitation from the plant side can arise in two opposite directions: either by defending chemically or physically against insects, or by evolving towards a cooperative interaction in which the cost of insect exploitation is balanced by an ecological service imposed by the plant (Dufaÿ and Anstett 2003). The latter case is notably encountered in mutualistic pollination systems, in which costs and benefits of plants and insects tend to be equilibrated. Ecological interactions thus include a strong coevolutionary component (Thompson 2009) that in cases of strict coevolution, leads to an increased specificity of the plant-insect relationship, or in cases of diffuse coevolution, maintains a generalist relationship (Lunau 2004).

Specific and obligate interactions, though much rarer than generalist relationships (Ollerton, Killick et al. 2007), represent simple cases for hypothesis testing and have therefore been frequently used as models by evolutionary biologists. Investigating biological systems with a reduced number of specifically interacting organisms allows understanding processes that would be much more difficult to disentangle when working with more relaxed interactions, as those happening in generalist relationships. Most studies emphasizing the fate of organisms in a context of strict coevolution and interested in unraveling evolutionary histories in pollination mutualisms have focused on large spatio-temporal scales, showing long-term processes putatively leading to cocladogenesis and codiversification (Agosta 2006), as for instance in the case of the fig and fig-wasps (Jousselin, van Noort et al. 2008). However, investigation at smaller evolutionary scales holds the potential to inform us on the origin and maintenance of specific mutualistic pollination. A few biological systems have

served for this purpose, as for instance, the interaction between *Trollius europaeus* L. (Ranunculaceae) and the pollinating flies within the genus *Chiastocheta* Pokorny (Diptera: Anthomyiidae), which represents a unique widespread European example of specific mutualism between plants and insects (Pellmyr 1989).

Trollius europaeus, the European globeflower, is a geophyte inhabiting cold meadows in the west-Palearctic, occupying southern European mountainous environments and being largely present in northern arctic-alpine areas of the continent (Pellmyr 1992). This species is characterized by a special flower morphology, adapted to the nursery pollination interaction it maintains with the *Chiastocheta* species complex (Pellmyr 1992). In this system, the plant is visited and specifically pollinated by the small Anthomyiids, which visit the flowers to lay eggs on carpels, since larvae feed specifically on *T. europaeus* seeds. There is strong evidence to consider that the mutualistic behaviour of *Chiastocheta* flies derives from phytophagy on different plant organs, since closely-related fly genera such as *Delia* Robineau-Desvoidy are generalist herbivores (Soós and Papp 1986). Considering this biological system in a coevolutionary framework, it is likely that herbivore pressure was released by evolving towards a specific pollination mechanism. While eight *Chiastocheta* species have been described based on taxonomic grounds (Michelsen 1985; Pellmyr 1992; Jaeger and Després 1998), a recent phylogenetic study by Espíndola et al. (unpublished data) has shown that the genetic identity of these species is far from being well circumscribed, with hybridization happening among several taxa. In fact, only three species are consistent from both genetic and morphological points of view: *C. rotundiventris, C. lophota* and *C. dentifera*; the remaining taxa appear either impossible to be identified by means other than morphology (*C. abruptiventris*), or extremely admixed with other morphotypes (*C. macropyga, C. trollii, C. inermella, C. setifera*). In that study, authors have also shown that diversification within the genus *Chiastocheta* arose during the Quaternary (up to 2.6 Mya). This specific interaction is thus much more recent than other systems such as those involving figs and fig wasps, which originated around 75 million years ago (Cruaud, A., University of Montpellier II, unpublished data).

Here, we exploit the unique attributes of this biological system to investigate the fate of specific mutualists at short-time evolutionary scales. We study the synchrony of diversification and more generally the congruence in phylogeographic histories of interacting species. This is done in order to address whether or not patterns similar to those found in more ancient specific mutualisms such as the case of figs and fig wasps (i.e., high level of congruence and large-scale co-cladogenesis) can be retrieved at smaller temporal scales. Because mutualistic species are strongly dependent on each other, we predict that *T. europaeus* and *Chiastocheta* flies should experience similar environmental and geographical effects on their dispersal and demographic components, and that their phylogeographic histories should thus be highly congruent. The alternative hypothesis is that the genetic structures of mutualistic species are different, what would indicate no effect of the interaction on the observed phylogeographic patterns.

The pollination mutualism involving *T. europaeus* and its pollinating flies is well-suited to test such a hypothesis. First, the phylogeography of European organisms is well understood. Two decades of phylogeographic studies in this continent have allowed unraveling general phylogeographic patterns and identifying spatial (Avise 2009) and ecological (Alvarez, Thiel-Egenter et al. 2009) factors responsible for the current spatial genetic structure of species. For instance, it is currently largely accepted that during glacial periods temperate species found refugia in southern peninsulas (Taberlet, Fumagalli et al. 1998; Hewitt 1999), and sometimes in northern refugia (Stewart and Lister 2001; Bhagwat and Willis 2008; Provan and Bennett 2008). Recently, Stewart, Lister et al. (2010) have summarized the effect of glacial and interglacial periods for cold-related species, which are found today under refugial situations occupying the southern European mountain ranges as well as northern areas. Second, the interaction between the globeflower and its pollinators has been widely studied at the European scale, especially in the Alps (Jaeger and Després 1998) and southern Scandinavia (Pellmyr 1989; Johannesen and Loeschcke 1996). In these studies, authors have identified a sound ecological background relative to the strength and characteristics of the interaction maintained by the different *Chiastocheta* species and to the dynamics of the

system. Third, new analytical approaches have been recently proposed in the field of phylogeography, to go further than simply qualitatively describing the distribution of lineages in space (Hickerson, Carstens et al. 2010). Probably the most ground-breaking of these recent approaches is the application of statistical phylogeography, which is derived from the theory of allele coalescence (Kingman 1982) and which proposes defining and statistically testing different phylogeographic scenarios (Knowles and Maddison 2002). This technique considers that the variation of coalescent times of trees contained in species (or population) trees can be used to create an expected distribution of a given statistic under a given scenario, which allows testing the probability of the empirical data to have been produced under that given scenario. The technique has been used for a little less than ten years, and it is currently acquiring importance, especially after the demonstration of its versatility when used in combination with ecological niche models and hindcasted spatial distributions (Carstens and Richards 2007), or with other geographical external data as glaciation and deglaciation times (Knowles 2001).

While one other study has focused on similar questions at small evolutionary scales using coalescent approaches (DeChaine and Martin 2006), our study is the first using a pollination mutualism to test for congruence in recent evolutionary histories and directly exploiting the known genetic structure of one of the organisms as a model into which the phylogeographic history of its partner should fit. Because the plant genetic structure has been recently shown to have been influenced by the Last Glacial Maximum (LGM, 21-18 thousand years ago - Kya; Espíndola, Pellissier et al. in preparation), we propose here to define and test hypotheses asking whether or not the phylogeographic patterns of the insects are driven by that of the plant. Based on three sequenced regions from the mitochondrial genome, we first infer the phylogeographies of each well-defined insect species (*i.e., C. rotundiventris, C. lophota* and *C. dentifera*) and further test if the phylogeographic patterns of the insects are more similar to a model fitting the plant's phylogeographic pattern than what would be expected by chance.

Material and Methods

Sampling and DNA amplification

Insect samples covering the whole European distribution range of the plant were collected during springs 2006-2008 and preserved in 70% ethanol. Samples were identified at the species level (Sup. Mat. 1) following Hennig (1976) and further confirmed by the European specialist of Anthomyiids, V. Michelsen (Natural History Museum of Denmark). In several cases, only expert identification allowed identifying morphotypes that were not described by Hennig (1976) (e.g., *C. lophota*). Only the three taxa presenting a consistent phylogenetic identity (*i.e., C. rotundiventris, C. dentifera,* and *C. lophota*; Espíndola, Buerki et al. in preparation) were further selected to infer phylogeographic patterns and demographic parameters.

DNA of 87 *C. rotundiventris*, 38 *C. dentifera* and 47 *C. lophota* samples was extracted using the QIAGEN DNeasy Animal Tissue extraction kit (QIAGEN, Hombrechtikon, Switzerland), following the manufacturer's indications. Three mitochondrial regions (*COI*, *COII* and *D-loop*) were amplified using primers shown in Table 1. PCRs were done in a 20µl mix consisting of 0.5X buffer, between 1 and 2.5 mM $MgCl_2$, 10mM dNTPs, 1 unit of GoTaq DNA polymerase (Promega, Dübendorf, Switzerland), 0.5 µM primers and 3 µl DNA and run in a TGradient thermocycler (Biometra, Goettingen, Germany). Two types of thermocycler programs were used, depending on the region amplified: for *COI* and *COII*, we used 1:30 min at 95°C, followed by 40 cycles of 35 secs at 95°C, 1 min at 52°C, 45 secs at 72°C, and a final elongation of 8 min at 72°C; for the *D-loop*, which is a highly A-T rich region, we used 5 min at 95°C, followed by 35 cycles of 1 min at 95°C, 1 min at 55°C, 2 min at 60°C and a final elongation of 5 min at 60°C. Amplified fragments were purified and sequenced at Macrogen Inc. (South Korea) and Fasteris S.A. (Switzerland). Chromatograms were corrected on ChromasPro 1.41 (Technelysium Pty. Ltd.), aligned either using a Clustal-Wallis algorithm

for *COI* and *COII* (Thompson, Higgins et al. 1994) as implemented in BioEdit 7.0.4.1 (Hall 1999) or applying the moderately accurate option proposed by the MAFFT v6 online alignment service for the *D-loop* (Katoh and Toh 2008). All alignments were further visually checked and corrected if necessary. Gaps were coded applying the simple gap-coding method of Simmons and Ochoterena (2000), as implemented in FastGap1.2 (Borchsenius 2009). Total, variable, constant and parsimony informative sites were calculated per species and mtDNA region using MEGA 4.0 (Tamura, Dudley et al. 2007).

Phylogenetic inferences and haplotype networks

MrBayes 3.1.2 (Ronquist and Huelsenbeck 2003) was used to infer Bayesian inference phylogenies for *C. dentifera* and *C. lophota* (no convergence was reached for *C. rotundiventris,* see Results). For each single species alignment, we parameterized two Markov-chain Monte Carlo (MCMC) runs with two chains, consisting of 50'000'000 generations, a temperature of 0.5 and one sampled tree every 1'000 generations. Data was partitioned for each mtDNA region using specific models of evolution following results obtained with MrAIC (Nylander 2004). A restriction model of evolution was assigned to the partition corresponding to gaps (coded as binary data). Convergence was assumed when standard variation between chains fell below 0.01, when the Potential Scale Reduction Factor index (Gelman and Rubin 1992) reached at most 1.002 , when a unimodal distribution of sampled parameters was retrieved and when sampling for all parameters presented values of ESS (Effective Sampling Size) higher than 200 (checked on Tracer v1.4, Rambaut and Drummond 2004). Half-compatible consensus trees were calculated applying burnins of 10'000'000 (*C. dentifera*) and 20'000'000 (*C. lophota*) generations.

Maximum Likelihood (ML) searches were done using RaxML 7.2.6 (Stamatakis 2006) with a 10'000 rapid bootstrap analyses followed by the search of the best-scoring ML tree in one single run. Here, we considered the three mtDNA regions as one single partition and did not account for gap information. A majority-rule consensus tree was further inferred from the

node information. This analysis was done using the facilities offered by the CIPRES portal (San Diego, CA, USA).

Maximum Parsimony (MP) searches were performed using parsimony ratchet (Nixon 1999) as implemented in PAUPrat (Sikes and Lewis 2001). Based on recommendations by Nixon (1999), ten independent searches were performed with 200 iterations and 15% of the parsimony informative characters perturbed. The shortest equally most parsimonious trees were combined to produce a majority rule consensus tree. To assess the support at each node, Bremer supports (Bremer 1988) were calculated using TreeRot 3.0 (Sorenson and Franzosa 2007).

All topologies were rooted based on the general relationships previously inferred and proposed by Després, Pettex et al. (2002) and Espíndola et al. (unpublished data), which considered the genus *Delia* Robineau-Desvoidy as the closest outgroup.

Finally, haplotype-networks were constructed using TCS 1.21 (Clement, Posada et al. 2000), applying a connectivity threshold of 95% (*i.e.,* networks were unlinked if they differed by more than 95% of their sequences). While gaps were considered as missing data for *C. lophota* and *C. rotundiventris*, they were considered as a fifth state in *C. dentifera,* because of the lower level of polymorphism in the latter species.

In order to identify the spatial genetic structure of each of the three species, supported clusters were plotted on maps using ArcMap (ESRI, Redlands CA, USA).

Phylogeographic hypothesis testing

To test whether the spatial genetic structures of the insects are congruent with that of the plant, we applied a statistical phylogeography approach, as proposed by Knowles and Maddison (2002). This approach consists in defining different alternative phylogeographic scenarios –represented by coalescent population trees– on which trees and data are simulated. Based on these simulations, a null distribution of a chosen statistic is produced and is used to test a given phylogeographic scenario based on empirical data. This is done

by evaluating empirical values in a percentage range (generally 2.5%-97.5%) of the null model distribution (Knowles and Maddison 2002). Because demographic parameters can have a non negligible influence on the time of coalescence (Maddison 1997), one of the key points when applying these methods is the use of values and parameters that specifically represent the empirical data. In order to apply these methods to our study we first defined different phylogeographic hypotheses. For this, we considered the spatial genetic structure unraveled in a previous study on the plant (Espíndola et al, submitted), which allowed us to rely on an external and independent source of clustering, a principle similar to the one Carstens and Richards (2007) applied when exploiting information from ecological-niche models. Specifically, population trees were defined by assigning insect samples based on the genetic cluster to which the plant populations had previously been assigned (see Figure 1). This assignment was straightforward since samples from insects and plants had been collected simultaneously in all localities visited. Because actors of this mutualism are cold-adapted species (*sensu* Stewart, Lister et al. 2010) and since the current distribution of the plant has been driven by the last range contraction following the LGM (in our case related to the termination of the last glacial period, Raymo 1997), we used splitting times of lineages as follows (Figure 1): in the working hypothesis H_1, the genetic clustering of the insect was hypothesized to have followed contraction of plant lineages after complete deglaciation of the main European lowlands (-13Kya, Raymo 1997), while in the alternative hypothesis H_2, the genetic clustering was hypothesized to have followed contraction of plant lineages immediately after the previous Glaxial Maximum (-23 Kya, Boulton, Dongelmans et al. 2004). The null hypotheses for each of these scenarios were that the splitting of insects' genetic clusters was not defined by the distribution of plant lineages (*i.e.,* phylogeographies of the insects and the plant were not related).

The demographic parameter Θ takes into account the effective population size (N_e) and the mutation rates (μ) ($\Theta = N_e * \mu$, in mitochondrial maternally inherited data; Beerli 2008) and was calculated for the two species in which a spatial genetic structure was retrieved (*i.e., C. lophota* and *C. dentifera*, see Results) using the MCMC approach applied in a Maximum-

Likelihood (ML) framework as implemented in Migrate-n (Beerli and Felsenstein 1999; Beerli and Felsenstein 2001) and run on the HPC-LSU cluster (Baton Rouge, LA, USA). Two runs of ten short and three long chains were used for 50,000,000 generations, sampling one value every 1,000 generations and applying a burnin period of 10'000'000 generations. Convergence between runs was verified using the Gelman criterion (Gelman and Rubin 1992) (option *gelman-convergence=YES:PAIRS*). Once this parameter was inferred, we applied it in the software ms (Hudson 2002) to simulate 1,000 tree topologies fitting the coalescent models represented by our hypotheses. Based on these topologies, sequences carrying the same characteristics as the empirical data (*i.e.,* sequence length, model of evolution, number of variable sites) were simulated for each simulated tree using seq-gen v1.3.2 (Rambaut and Grassly 1997). Following the approach proposed by Carstens et al. (2005), we considered two rates of mutation (μ): either $1x10^{-7}$ or $1x10^{-8}$. Using these sequences, parsimony heuristic searches were performed with PAUP*4b10 (Swofford 2003), with TBR branch swapping , maxtrees = 100 and 10 random-addition sequence replications. A strict consensus topology of the most parsimonious trees was computed for each dataset and the resulting trees (1'000) were used to create a null distribution of the test statistic.

To estimate the fit or discordance between gene (*i.e.*, simulated and empirical data) and species (*i.e.*, phylogeographic scenarios) trees (Knowles and Maddison 2002), we used the statistic S of Slatkin & Maddison (1989), which indicates the minimum number of migration events (parsimony steps) required to fit the phylogeny into the population tree. This statistic is optimal in our case, because we aim at testing the congruence between the plant and the insects' spatial genetic structures. Because our hypotheses consider that dispersal events should have happened in parallel among the different actors involved, using a measure of migration appears appropriate. Calculations of S for simulated trees and empirical data (i.e., the Bayesian phylogenetic tree inferred from real data) were performed using MESQUITE 2.72 (Maddison and Maddison 2009). We considered rejection of hypotheses H_1 or H_2 if the empirical value fell outside the 2.5%-97.5% range of the simulated data (i.e., $\alpha = 5\%$).

278

279 **Results**

280 The geographic distribution of the sampled species covered their entire known European
281 distribution ranges (Figures 2, 3 and 4) (Pellmyr 1992). Total length, number of constant,
282 variable, parsimony informative sites and number of gaps per species are shown for each
283 mtDNA region in Table 2. The *D-loop* region was by far the most variable and its
284 corresponding alignment was the only one comprising gaps (gaps were absent in both *COI*
285 and *COII* alignments).

286 ***Phylogeographic inferences***

287 *C. rotundiventris*

288 No convincing topology could be retrieved using the Bayesian inference approach:
289 independent runs did not converge, even though trials with different MCMC parameters,
290 models of evolution, and among-partition linkage were performed. As a consequence,
291 Bayesian topologies were not taken into consideration for *C. rotundiventris*. In contrast, ML
292 and MP approaches provided partially congruent results, indicating that resolution is very low
293 in this taxon (Figure 2a), with extremely short branch lengths (ML) or polytomized tips (MP),
294 though about 5% of the nucleotides among the three regions were variable. When analyzing
295 patterns of the haplotype network inferred for *C. rotundiventris*, some spatial structure could,
296 however, be recovered among the 44 haplotypes (of which eight were shared by several
297 samples and 36 were private to single samples) (Figure 2b): the most abundant haplotype
298 (red) was absent from Scandinavia, while the purple and yellow haplotypes were restricted
299 to this region. Other haplotypes were private to some populations, as the grey (south
300 Scandinavia), the blue (northern Poland) and the black (Balkans).

301 *C. lophota*

Unlike *C. rotundiventris*, Bayesian inference runs converged well, and all methods (Bayesian, ML and MP) were fully congruent in terms of both topology and support of the three main clades (Figure 3a). The same three groups were also retrieved in the haplotype network analysis (Figure 3b), and relationships between clades were also congruent with the topological analysis (i.e., haplotypes from blue and red groups presented closer relationships -a maximum of three steps- than those relating them with the yellow clade). The distribution of clades indicated a clear, geographically structured genetic distribution (Figure 3c). While the yellow clade was restricted to eastern populations, the red clade was exclusive of the southwestern Alps and the blue clade was present mainly in the western distribution and on the northern edge of the Alps.

C. dentifera

The three phylogenetic approaches yielded similar topologies and node supports, except in the MP approach, which harbored low Bremer supports (Figure 4a). Analyses retrieved four main clades (blue, red, yellow and purple) supported by at least one method, and two additional clans (*sensu* Wilkinson, McInerney et al. 2007) placed in polytomies (green and orange). The haplotype network approach indicated the presence of three shared and 15 unique haplotypes (Figure 4b). Haplotypes were tightly interrelated, with a maximum of five steps between the two more distant haplotypes. Some of the shared haplotypes matched with clades defined by the phylogenetic approach (*i.e.*, red clade), while others contained only some part of the samples included in the clades (*i.e.*, green and yellow). The distribution of clades indicated a trend towards spatial genetic structure, with samples from several clades restricted to particular regions (Figure 4c): the yellow clade spread over southern Scandinavian, as well as several locations in the Alps; the blue and red clades were restricted to Scandinavia, while the orange clan and purple clade were mainly present in Alpine locations. The remaining green clan appeared present in all regions, but was the only one observed in the Pyrenees and in the northeastern most populations.

Phylogeographic hypothesis testing

Testing phylogeographic scenarios in fly species using the canvas of the plant genetic structure could be done only in species for which Bayesian inference analyses converged and yielded supported clades (*i.e., C. dentifera* and *C. lophota*). Maximum Likelihood Estimates (MLE) of theta values for *C. dentifera* and *C. lophota* were 0.000229 and 0.001486, respectively, providing N_e equal to 26'220 ($\mu = 1x10^{-7}$) and 262'200 ($\mu = 1x10^{-8}$) for *C. dentifera* and to 14'860 ($\mu = 1x10^{-7}$) and 148'600 ($\mu = 1x10^{-8}$) for *C. lophota*. Calculated S values for the empirical data were 24, for *C. dentifera* and 27 for *C. lophota*.

Phylogeographic scenario-testing indicated that while we could not statistically reject the null hypothesis for *C. lophota*, at least when considering a mutation rate equal to $1x10^{-7}$ ($p_{H0-H1}=0.39$ and $p_{H0-H2}=0.55$; Figure 5), this was not the case for *C. dentifera*, for which the calculated S value fell completely within the simulated distributions under the assumption of a phylogeographic structure based on the plant genetic structure, whatever mutation rate was considered [$p_{H1}=0.76$ ($\mu = 1x10^{-7}$) or $p_{H1}=0.39$ ($\mu = 1x10^{-8}$) and $p_{H2}=0.77$ ($\mu = 1x10^{-7}$) or $p_{H2}=0.52$ ($\mu = 1x10^{-8}$); Figure 5].

Discussion

Though the phylogeography of European organisms has been revisited several times in the last two decades (e.g., Taberlet, Fumagalli et al. 1998; Hewitt 1999; Schmitt 2009), no studies have treated the fate of organisms involved in specific and obligate interactions in a framework of comparative phylogeography. In contrast to our hypothesis of congruence among spatial genetic structures of the different actors of a plant-pollinator mutualism, the three pollinator species studied here show different phylogeographic patterns, which are only in part congruent with patterns found in the plant. While *Chiastocheta rotundiventris* had no clear, geographically consistent genetic structure (haplotypes and topologies were not congruent; Figure 2), *C. lophota* (Figure 3) and *C. dentifera* (Figure 4) presented well-defined spatial genetic structures whatever method was used. This result is a first evidence

of the overall incongruence in the spatial genetic structure of the partners of this interaction, despite their closely related ecologies.

Chiastocheta rotundiventris is a species that has long been considered as the "most mutualistic" of all those interacting with *Trollius europaeus* (Pellmyr 1992). First, it has been shown to visit flowers at the very beginning of the flowering period, thus being present when the flower is extremely fresh and receptive (Pellmyr 1989) and laying only one egg per flower head, meaning a lower price to be paid by the flower in terms of number of larvae to host (Pompanon, Pettex et al. 2006). Among the three studied pollinators, *C. rotundiventris* appears as the species with the lower spatial genetic structure, indicating high levels of admixture between populations (Figure 2). This could be explained by long-distance dispersal abilities of this species, or by a scenario of incomplete lineage sorting in which colonization of the current distribution of this fly originated from one or a few neighboring regions during glacial periods. Concerning the first idea, it is known that small insects can sometimes exploit air currents to travel extremely long distances (Compton 2001); the likelihood of this hypothesis could even be enhanced by the mountainous habitat of these flies (in the southern part of their distribution at least), more prone to experience strong winds. In Europe, an absence of spatial genetic structure has been also unraveled in other cold-adapted organisms such as *Ligusticum mutellinoides* (Alvarez, Thiel-Egenter et al. 2009) or *Ranunculus pygmaeus* (Schönswetter, Popp et al. 2006).

However, if we consider the haplotypic approach applied in this study some indications show that even though high admixture (*i.e.*, haplotypes are extremely interconnected) is present in *C. rotundiventris*, some geographic grouping is possible (Figure 2b). The distribution of these haplotypes sometimes presents some geographic structure, as it is the case of the central European/South-Scandinavian yellow haplotype, or the isolated Balkan black type. Probably this last case is the most clearly structured, since it is the only clade which is both supported in the phylogenetic and haplotypic approaches. Finally, despite

being somehow geographically structured, all the remaining haplotypes are separated by very small numbers of steps.

The incongruence between the haplotypic and topological approaches applied for *C. rotundiventris* indicates that it is possible that some past geographic structure was present, but that it is currently being blurred by interpopulational exchanges. It is also interesting to note that even though it is currently impossible to test phylogeographic hypotheses based on the topology obtained in this study, the distributions of some of these haplotypes (*i.e.*, yellow, purple, black, red) are similar distribution to that of the host plant (Espíndola et al, unpublished data). In the future, it would be worth testing more variable molecular markers to more deeply investigate the genetic structure of this species. An approach based on next-generation sequencing technologies might be a good option, as soon as this technology will become available at a reasonable cost to phylogeographers.

Unlike *C. rotundiventris*, *C. lophota* shows a phylogeographic pattern with a clear and supported geographic structure (Figure 3c), with all methods yielding congruent results. It is thus probable that the different clades have been isolated for a long time, and that their fragmented distributions started to come into contact only recently. This hypothesis is suggested by the long branches separating eastern (yellow) and western (blue and red) genetic groups (Figure 3a). This genetic structure has also been previously observed in other arctic-alpine European organisms, as the butterfly *Erebia epiphron* (Schmitt, Hewitt et al. 2006) or the gentian *Comastoma tenellum* (Schönswetter, Tribsch et al. 2004), which also appeared structured with east, central and western genetic lineages, coming into contact in the Alpine range. It is possible that this genetic characteristic is a consequence of the presence of this species in mountainous environments (*C. lophota* presents a distribution range which occupies only the southern European mountains), which in southern Europe represent massifs isolated by large areas of low elevation plains, which could have enhanced their isolation during long periods of time.

Finally, the spatial genetic structure of *C. dentifera* indicates some geographic structuring (Figure 4c), especially in Scandinavia (red and blue groups), and part of the southeastern (Tatra) distribution (purple). However, similarly to *C. rotundiventris*, some lineages are widespread, as the green clan, which is present in all regions. Like *C. rotundiventris*, but to a lesser extent, haplotypes appear interconnected, probably indicating current or recent genetic exchange between populations.

If similarities had to be identified between the three species, we could mention that *C. rotundiventris* and *C. dentifera* present –despite the reduced genetic structure of the first- specific lineages which appear restricted to Northern Scandinavia (yellow in *C. rotundiventris*, red in *C. dentifera*, Figures 2 and 4). Moreover, both harbor one genetic group absent of northern Scandinavia (red in *C. rotundiventris*, yellow in *C. dentifera*; Figures 2 and 4). Because *C. lophota* presented a clear, geographically structured genetic pattern, it is difficult to identify any parallel between this species and the other two, which did not present such a structure. It appears, nonetheless, that in all cases, the Alps acted as a center of current genetic diversity, which can be explained either by *in situ* glacial survival (Brochmann, Gabrielsen et al. 2003) or by postglacial colonization during demographical contractions, as proposed by the paradigm of the multiple recolonization of the Alps (Schönswetter, Stehlik et al. 2005).

Because the three species present very different spatial genetic structures, despite sharing medium to high levels of mtDNA genetic diversity, it seems that they have endured distinct demographic and dispersal processes. As a consequence, we argue that the phylogeography of the plant does not seem to present congruencies with those of all fly species with which it is ecologically associated.

The effect of the plant on the genetic structure of the mutualistic flies

Statistical phylogeography is a relatively new field that provides the possibility to quantitatively test phylogeographic hypotheses (Knowles and Maddison 2002). It is thus an

ideal tool to identify drivers of phylogeographic patterns, and in our case, it fits well our questions related to congruence in the spatial genetic structures of interacting species. Despite the fact that the different insect species are tightly related to the plant from an ecological point of view, our results show that there is a high heterogeneity in the level of congruence between each insect species and the plant spatial genetic structures. Hypothesis testing revealed that while one species (*C. lophota*) does not present a genetic structure likely to be driven by that of the plant (H_0 not rejected, Figure 5), this is not the case of *C. dentifera*. This last species presents a genetic pattern that does not significantly differ of what would be expected in a case of plant-insect phylogeographic congruence (both H_1 and H_2 could not be rejected, while the two null hypotheses were rejected; Figure 5).

In contrast to *C. lophota*, which presents a genetic structure that is largely explained by geographic factors (Figure 2) and that is not explained by the genetic structure of the plant, *C. dentifera* seems more tightly related to the demographic and dispersal dynamics of *T. europeaus*. Considering this, it is interesting to note that our test could not differentially reject any of the non null-hypotheses for *C. dentifera*, indicating either that the plant has probably been shaping the genetic structure of the insect for longer than one glacial period. If this is true, it is possible that the insect and the plant depend strongly on each other to assure survival even under environmental changes, and that demographic events and variations occur in a parallel manner for these species. This should however be tested in the future, considering that because the plant interacts with several fly species, it is possible that it is able to move more freely than the insects.

Coupling phylogeographic patterns of organisms that are ecologically associated is a recent practice in the field of phylogeography. In this context, the case of *Fagus* and *Epifagus* parasites is one of the unique examples in which a relationship between interacting species has been demonstrated at a large scale (Tsai and Manos 2010). The present study provides additional support to the idea that recent histories of interacting organisms are

partly related, confirming that some processes occurring at large spatio-temporal scales (e.g., synchronous co-cladogenesis) are also relevant at smaller evolutionary scales.

However, the pollination system studied here informs on the heterogeneity of scenario matching in species involved in tightly-related mutualistic interactions, demonstrating that it is important that each interacting organism studied is considered independently of the others. The example of *T. europaeus* and its specific pollinators is thus highly instructive since in this case, despite closely related ecologies, flies present different demographic and dispersal histories. It is however possible that because the niches of these species strongly overlap, interspecific competition is high, and differentiation in terms of population dynamics, dispersion capabilities and population sizes could thus be enhanced. This idea of escaping from interspecific competition in the *Chiastocheta* species complex has also been demonstrated in other ecologically focused studies, which have shown that different taxa display differential oviposition patterns (Després and Jaeger 1999) and seed-feeding behaviors (Pompanon, Pettex et al. 2006). It is thus possible that as one of the most basal species in the group (Després, Pettex et al. 2002), *C. dentifera* still harbors demographic and dispersal patterns similar to that of the plant, whereas derived taxa such as *C. lophota* have experienced niche switches towards leading to different phylogeographic patterns.

Conclusion

In the present study we show that interacting species may not only present different phylogeographic patterns, but also that forces shaping phylogeographic structures may differ among species: the presence of the plant during distributional changes for *C. dentifera*, geography for *C. lophota*, and eventually migration capabilities and geography for *C. rotundiventris*.

In the future it would be worth testing these hypotheses on the plant genetic material, based on the insects' genetic structures. Indeed, it is plausible that in the same way that we could not reject the hypothesis of a genetic structure driven by the plant for *C. dentifera*, the

485 genetic structure of the plant could also be influenced by one or another fly species.
486 However, in order to do this, analytical bases should be first firmly established before
487 allowing us to test hypotheses based on dominant data issued of genomic fingerprinting (i.e.,
488 the spatial genetic structure of *T. europaeus* was retrieved using AFLP).

489 **Acknowledgments**

490 The authors would like to thank the invaluable help of Dr. Y. Triponez, N. Magrou, T.
491 Suchan, Dr. P. Lazarevic, D. Gyurova, R. Lavigne, L. Juillerat, N. Villard and R. Arnoux
492 during field work. The informatics support provided by A. Jundt from LSU HPC was
493 extremely appreciated during analyses. This work was funded by the grant “Fonds des
494 Donations” of the University of Neuchâtel attributed to AE and by the Swiss National Science
495 Foundation (project No. 3100A0-116778).

496 **References**

497 Agosta, S. J. (2006). "On ecological fitting, plant-insect associations, herbivore host shifts,
498 and host plant selection." OIKOS **114**(3): 556-565.
499 Alvarez, N., C. Thiel-Egenter, et al. (2009). "History or ecology? Substrate type as a major
500 driver of spatial genetic structure in Alpine plants." Ecology Letters **12**(7): 632-640.
501 Avise, J. C. (2009). "Phylogeography: retrospect and prospect." Journal of Biogeography **36**:
502 3-15.
503 Beerli, P. (2008). Migrate documentation. Tallahassee FL, USA, Florida State University**:** 97.
504 Beerli, P. and J. Felsenstein (1999). "Maximum-likelihood estimation of migration rates and
505 effective population numbers in two populations using a coalescent approach."
506 Genetics **152**: 763-773.
507 Beerli, P. and J. Felsenstein (2001). "Maximum likelihood estimation of a migration matrix
508 and effective population sizes in n subpopulations by using a coalescent approach."
509 Proceedings of the National Academy of Sciences of the United States of America
510 **98**(8): 4563-4568.
511 Bhagwat, S. A. and K. J. Willis (2008). "Species persistence in northerly glacial refugia of
512 Europe: a matter of chance or biogeographical traits?" Journal of Biogeography **35**:
513 464-482.

514 Borchsenius, F. (2009). FastGap 1.2. Department of Biological Sciences, University of
515 Aarhus, Denmark, Published online at
516 http://192.38.46.42/aubot/fb/FastGap_home.htm.
517 Boulton, G. S., P. Dongelmans, et al. (2004). Evidence of European ice sheet fluctuation
518 during the last glacial cycle. Quaternary Glaciations - Extent and Chronology - Part I:
519 Europe. E. J. and P. L. Gibbard. Amsterdam, the Netherlands**:** 441-460.
520 Bremer, K. (1988). "The limits of amino acid sequence data in angiosperm phylogenetic
521 reconstruction." Evolution **42**: 795-803.
522 Brochmann, C., T. M. Gabrielsen, et al. (2003). "Glacial survival or *tabula rasa*? The history
523 of North Atlantic biota revisited." Taxon **52**: 417-450.
524 Carstens, B. C., S. J. Brunsfeld, et al. (2005). "Investigating the evolutionary history of the
525 Pacific Northwest mesic forest ecosystem: hypothesis testing within a comparative
526 phylogeographic framework." Evolution **59**(8): 1639-1652.
527 Carstens, B. C. and C. L. Richards (2007). "Integrating coalescent and ecological niche
528 modeling in comparative phylogeography." Evolution **61**(6): 1439-1454.
529 Clement, M., D. Posada, et al. (2000). "TCS: a computer program to estimate gene
530 genealogies." Molecular Ecology **9**(10): 1657-1660.
531 Compton, S. G. (2001). Sailing with the wind: dispersal by small flying insects. Dispersal
532 Ecology: the 42nd Symposium of the British Ecological Society, University of
533 Reading, Blackwell Science, Oxford, UK.
534 DeChaine, E. G. and A. P. Martin (2006). "Using coalescent simulations to test the impact of
535 quaternary climate cycles on divergence in an alpine plant-insect association."
536 Evolution **60**(5): 1004-1013.
537 Després, L. and N. Jaeger (1999). "Evolution of oviposition strategies and speciation in the
538 globeflower flies *Chiastocheta* spp. (Anthomyiidae)." Journal of Evolutionary Biology
539 **12**: 822-831.
540 Després, L., E. Pettex, et al. (2002). "Speciation in the globeflower fly *Chiastocheta* spp.
541 (Diptera: Anthomyiidae) in relation to host plant species, biogeography, and
542 morphology." Molecular Phylogenetics and Evolution **22**(2): 258-68.
543 Dufaÿ, M. and M.-C. Anstett (2003). "Conflicts between plants and pollinators that reproduce
544 within inflorescences: evolutionary variations on a theme." OIKOS **100**: 3-14.
545 Ehrlich, P. R. and P. H. Raven (1964). "Butterflies and Plants: A study in coevolution."
546 Evolution **18**(4): 586-607.
547 Espíndola, A., S. Buerki, et al. (in preparation). "Evolutionary history of Chiastocheta species
548 interacting with *Trollius europaeus*."

549 Espíndola, A., L. Pellissier, et al. (in preparation). "No fossils, no answers? Inferring
550 postglacial genetic consequences in cold-adapted species when neither ancient DNA
551 nor macroremains are available." Journal of Biogeography.
552 Futuyma, D. J. and A. A. Agrawal (2009). "Macroevolution and the biological diversity of
553 plants and herbivores." Proceedings of the National Academy of Sciences of the
554 United States of America **106**(43): 18054-18061.
555 Gelman, A. and D. B. Rubin (1992). "Inference from iterative simulation using multiple
556 sequences." Statistical Science **7**: 457-472.
557 Hall, T. A. (1999). "BioEdit: a user-friendly biological sequence alignment editor and analysis
558 program for Windows 95/98/NT." Nucleic Acids Symposium Series **41**: 95-98.
559 Hennig, W., Ed. (1976). Anthomyiidae. Die Fliegen der Palaearktischen Region. Stuttgart, E.
560 Schweizerbart.
561 Hewitt, G. M. (1999). "Post-glacial re-colonization of European biota." Biological Journal of
562 the Linnean Society **68**(1-2): 87-112.
563 Hickerson, M. J., B. C. Carstens, et al. (2010). "Phylogeography's past, present, and future:
564 10 years after Avise, 2000." Molecular Phylogenetics and Evolution **54**(1): 291-301.
565 Hudson, R. R. (2002). "Generating samples under a Wright-Fisher neutral model."
566 Bioinformatics **18**: 337-338.
567 Jaeger, N. and L. Després (1998). "Obligate mutualism between *Trollius europaeus* and its
568 seed-parasite pollinators *Chiastocheta* flies in the Alps." Comptes Rendus De
569 L'Académie Des Sciences Serie III-Sciences De La Vie-Life Sciences **321**(9): 789-
570 796.
571 Johannesen, J. and V. Loeschcke (1996). "Distribution, abundance and oviposition patterns
572 of four coexisting *Chiastocheta* species (Diptera: Anthomyiidae)." Journal of Animal
573 Ecology **65**: 567-576.
574 Jousselin, E., S. van Noort, et al. (2008). "One fig to bind them all: host conservatism in a fig
575 wasp community unraveled by cospeciation analyses among pollinating and
576 nonpollinating fig wasps." Evolution **62**(7): 1777-1797.
577 Katoh, K. and H. Toh (2008). "Recent developments in the MAFFT multiple sequence
578 alignment program." Briefings in Bioinformatics **9**(4): 286-298.
579 Kingman, J. F. C. (1982). "The coalescent." Stochastic Processes and Their Applications **13**:
580 235-248.
581 Knowles, L. L. (2001). "Did the pleistocene glaciations promote divergence? Tests of explicit
582 refugial models in montane grasshopprers." Molecular Ecology **10**(3): 691-701.
583 Knowles, L. L. and W. P. Maddison (2002). "Statistical phylogeography." Molecular Ecology
584 **11**(12): 2623-2635.

585 Lunau, K. (2004). "Adaptive radiation and coevolution - pollination biology case studies."
586 Organisms Diversity & Evolution **4**(3): 207-224.
587 Maddison, W. P. (1997). "Gene trees in species trees." Systematic Biology **46**(3): 523-536.
588 Maddison, W. P. and D. R. Maddison (2009). Mesquite: a modular system for evolutionary
589 analysis.
590 Michelsen, V. (1985). "A revision of the Anthomyiidae (Diptera) described by J. W.
591 Zetterstedt." Steenstrupia **11**(2): 37-65.
592 Nixon, K. C. (1999). "The parsimony ratchet, a new method for rapid parsimony analysis."
593 Cladistics **15**: 407-414.
594 Nylander, J. A. A. (2004). Mraic.pl. Uppsala, Sweden, Distributed by the author.
595 Ollerton, J., A. Killick, et al. (2007). "Multiple meanings and modes: on the many ways to be
596 a generalist flower." Taxon **56**(3): 1-12.
597 Pellmyr, O. (1989). "The cost of mutualism - interactions between *Trollius europaeus* and its
598 pollinating parasites." Oecologia **78**(1): 53-59.
599 Pellmyr, O. (1992). "The phylogeny of a mutualism: evolution and coadaptation between
600 *Trollius* and its seed-parasitic pollinators." Biological Journal of the Linnean Society
601 **47**: 337-365.
602 Pompanon, F., E. Pettex, et al. (2006). "Patterns of resource exploitation in four coexisting
603 globeflower fly species (*Chiastocheta* sp.)." Acta Oecologica **29**(2): 233-240.
604 Provan, J. and K. D. Bennett (2008). "Phylogeographic insights into cryptic glacial refugia."
605 Trends in Ecology & Evolution **23**(10): 564-571.
606 Rambaut, A. and A. J. Drummond (2004). Tracer v1.4. Edinburgh.
607 Rambaut, A. and N. C. Grassly (1997). "Seq-Gen: An application for the Monte Carlo
608 simulation of DNA sequence evolution along phylogenetic trees." Computational and
609 Applied Biosciences **13**: 235-238.
610 Raymo, M. E. (1997). "The timing of major climate terminations." Paleoceanography **12**(4):
611 577-585.
612 Ronquist, F. and J. P. Huelsenbeck (2003). "MrBayes 3: Bayesian phylogenetic inference
613 under mixed models." Bioinformatics **19**(12): 1572-1574.
614 Schluter, D. (2000). The ecology of adaptive radiation. Oxford, UK, Oxford University Press.
615 Schmitt, T. (2009). "Biogeographical and evolutionary importance of the European high
616 mountain systems." Frontiers in Zoology **6**: 9.
617 Schmitt, T., G. M. Hewitt, et al. (2006). "Disjunct distributions during glacial and interglacial
618 periods in mountain butterflies: *Erebia epiphron* as an example." Journal of
619 Evolutionary Biology **19**(1): 108-13.

620 Schönswetter, P., M. Popp, et al. (2006). "Rare arctic-alpine plants of the European Alps
621 have different immigration histories: the snow bed species *Minuartia biflora* and
622 *Ranunculus pygmaeus*." Molecular Ecology **15**(3): 709-720.
623 Schönswetter, P., I. Stehlik, et al. (2005). "Molecular evidence for glacial refugia of mountain
624 plants in the European Alps." Molecular Ecology **14**(11): 3547-3555.
625 Schönswetter, P., A. Tribsch, et al. (2004). "Amplified fragment length polymorphism (AFLP)
626 suggests old and recent immigration into the Alps by the arctic-alpine annual
627 *Comastoma tenellum* (Gentianaceae)." Journal of Biogeography **31**: 1673-1681.
628 Sikes, D. S. and P. O. Lewis (2001). PAUPRat, Beta software, version 1: PAUP
629 implementation of the parsimony ratchet, Distributed by the authors.
630 Simmons, M. P. and H. Ochoterena (2000). "Gaps as characters in sequence-based
631 phylogenetic analyses." Systematic Biology **49**(2): 369-381.
632 Simpson, G. G. (1953). The major features of evolution. New York, USA, Columbia
633 University Press.
634 Soós, Á. and L. Papp (1986). Catalogue of Palaearctic Diptera. Budapest, Hungarian Natural
635 History Museum.
636 Sorenson, M. D. and E. A. Franzosa (2007). Treerot, version 3. Boston MA, USA, Distributed
637 by the authors.
638 Stamatakis, A. (2006). "RAxML-VI-HPC: Maximum Likelihood-based phylogenetic analyses
639 with thousands of taxa and mixed models." Bioinformatics **22**(21): 2688-2690.
640 Stewart, J. R. and A. M. Lister (2001). "Cryptic northern refugia and the origins of the
641 modern biota." Trends in Ecology & Evolution **16**(11): 608-613.
642 Stewart, J. R., A. M. Lister, et al. (2010). "Refugia revisited: individualistic responses of
643 species in space and time." Proceedings of the Royal Society of London B **277**: 661-
644 671.
645 Swofford, D. L. (2003). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other
646 Methods). Version 4. Sunderland, Massachusetts, Sinauer.
647 Taberlet, P., L. Fumagalli, et al. (1998). "Comparative phylogeography and postglacial
648 colonization routes in Europe." Molecular Ecology **7**(4): 453-464.
649 Tamura, K., J. Dudley, et al. (2007). "MEGA4: Molecular Evolutionary Genetics Analysis
650 (MEGA) software version 4.0." Molecular Biology and Evolution **24**: 1596-1599.
651 Thompson, J. D., D. G. Higgins, et al. (1994). "CLUSTAL W: improving the sensitivity of
652 progressive multiple sequence alignment through sequence weighting, position-
653 specific gap penalties and weight matrix choice." Nucleic Acids Res **22**(22): 4673-
654 4680.
655 Thompson, J. N. (2009). "The coevolving web of life." American Naturalist **173**(2): 125-140.

Tsai, Y.-H. E. and P. S. Manos (2010). "Host density drives the postglacial migration of the tree parasite, *Epifagus virginiana*." Proceedings of the National Academy of Sciences of the United States of America **107**(39): 17035-17040.

Wilkinson, M., J. O. McInerney, et al. (2007). "Of clades and clans: terms for phylogenetic relationships in unrooted trees." Trends in Ecology & Evolution **22**(3): 114-115.

663 Table 1 – Names, sequences, annealing temperatures and references of primers used for

664 sequencing mitochondrial region sin *Chiastocheta spp.*

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	52	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	52	Després and Jaeger, 1999
D-loop	TM-N-193	TGG GGT ATG AAC CCA GTA GC	55	Simon et al. 1994
	SR-J-14612	AGG GTA TCT AAT CCT AGT TT	55	Simon et al. 1994

665

666

667 Table 2 – Number of total, constant, variable and parsimony informative sites (PI), as well
668 as number of identified gaps per amplified region and species in *Chiastocheta*.

Species	Region	Sites (base-pairs)				
C. dentifera		Constant	Variable	PI	Total	Gaps
	COI	653	4	1	657	0
	COII	475	4	0	479	0
	D-loop	883	32	15	1168	75
	Total	**2011**	**40**	**16**	**2304**	**75**
C. lophota						
	COI	644	13	6	657	0
	COII	470	9	6	479	0
	D-loop	768	60	29	855	62
	Total	**1882**	**82**	**41**	**1991**	**62**
C. rotundiventris						
	COI	645	12	4	657	0
	COII	457	22	5	479	0
	D-loop	865	56	27	1168	87
	Total	**1967**	**90**	**36**	**2304**	**87**

669

670

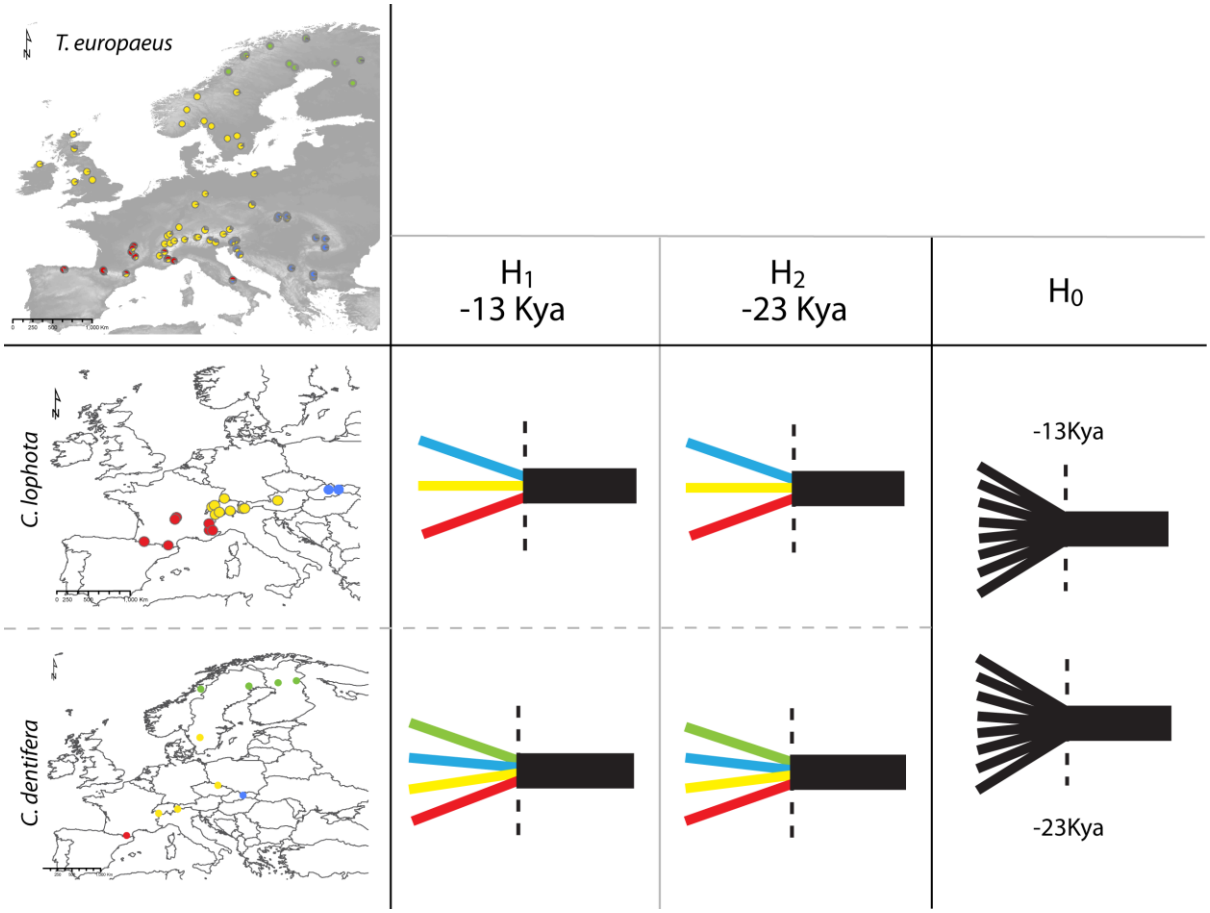


Figure 1 –Phylogeographic scenarios to be tested by coalescent-based phylogeographic statistics, for *Chiastocheta lophota* and *C. dentifera*, considering the inferred genetic identity of populations for *T. europaeus* based on Espíndola et al. submitted. Maps for *C. lophota* and *C. dentifera* show the position of sampled populations and the expected genetic identity they should present if the genetic structure of the fly was congruent to that of the plant (map on top). Fly populations included in each population tree branch are identified by the same colors on the map and the phylogeographic scenarios. H_1= Phylogeographic scenario 1: splitting occurred at the termination of the LGM (-13 Kya). H_2= Phylogeographic scenario 2: splitting occurred at the previous termination (-23 Kya). H_0= Null hypotheses for each of the two first phylogeographic scenarios.

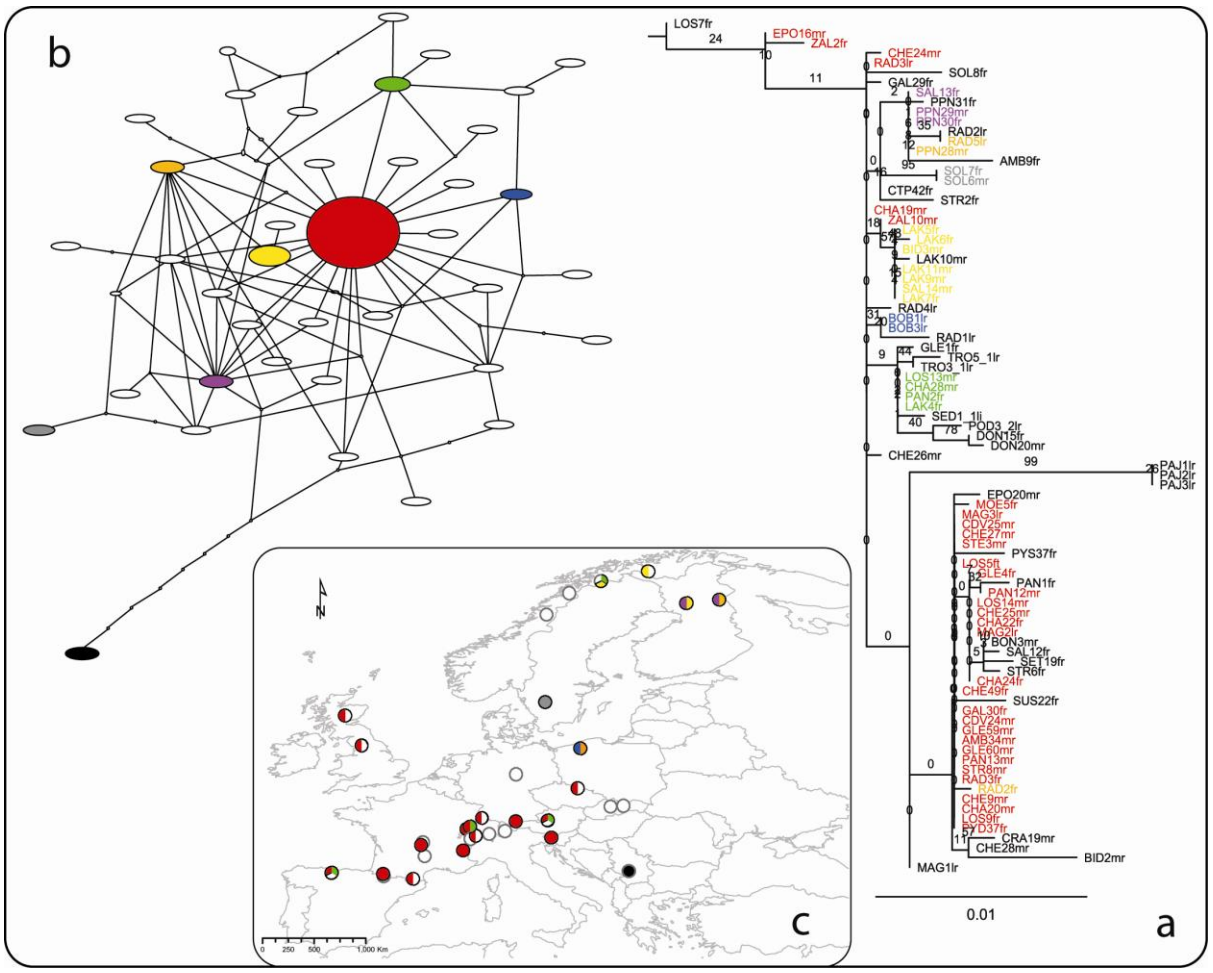


Figure 2 – Molecular inferences (a and b) and geographic distribution of haplotypes (c) in *Chiastocheta rotundiventris*. a- Maximum Likelihood (ML) phylogeny. Values on branches indicate ML supports considering 10,000 bootstraps. Colors indicate haplotypes (see b). b- Haplotype network. Colors show shared haplotypes; empty circles indicate unique haplotypes. c- Geographic distribution of haplotypes. Colors correspond to b. Sites where several colors appear indicate locations comprising different haplotypes.

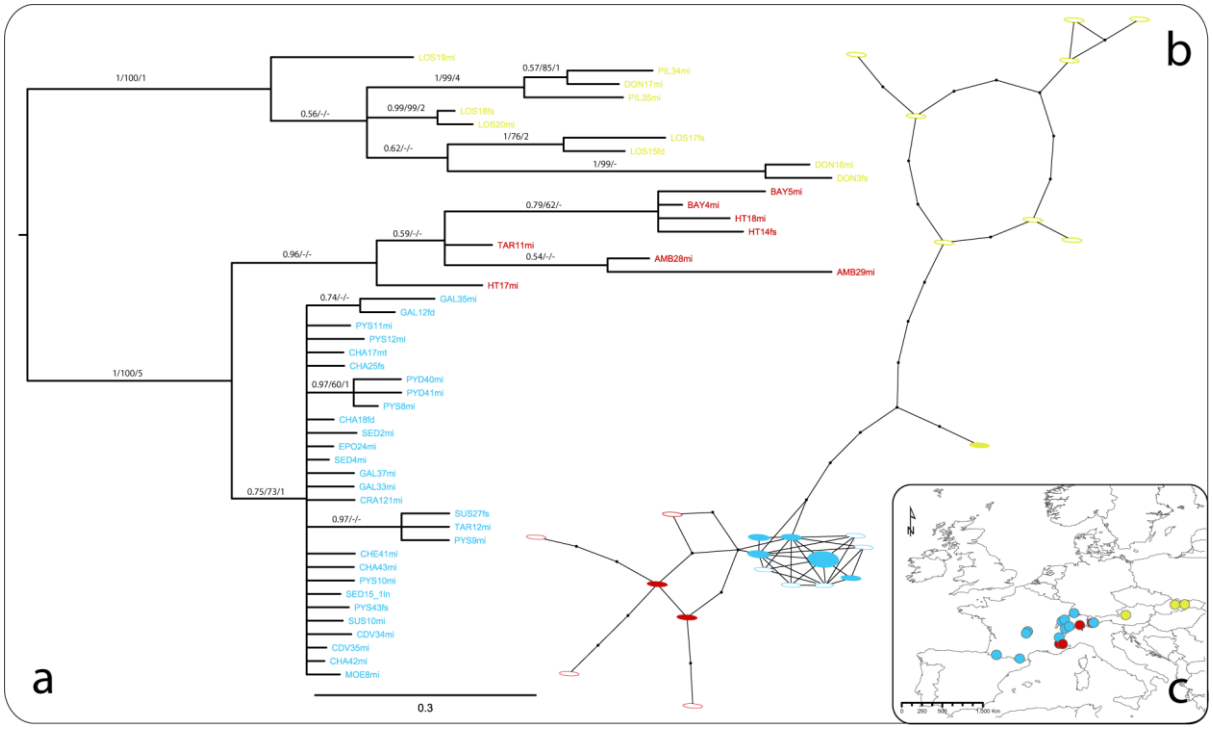


Figure 3 – Molecular inferences (a and b) and geographic distribution of haplotypes (c) in *Chiastocheta lophota*. a- Bayesian topological inference. Values on branches indicate Bayesian supports, Maximum Likelihood supports and Maximum Parsimony Bremer supports. Colors indicate main clades. b- Haplotype network. Colors indicate clades identified in a. Empty circles indicate unique haplotypes; filled circles indicate shared haplotypes. c- Geographic distribution of main clades. Colors correspond to a.

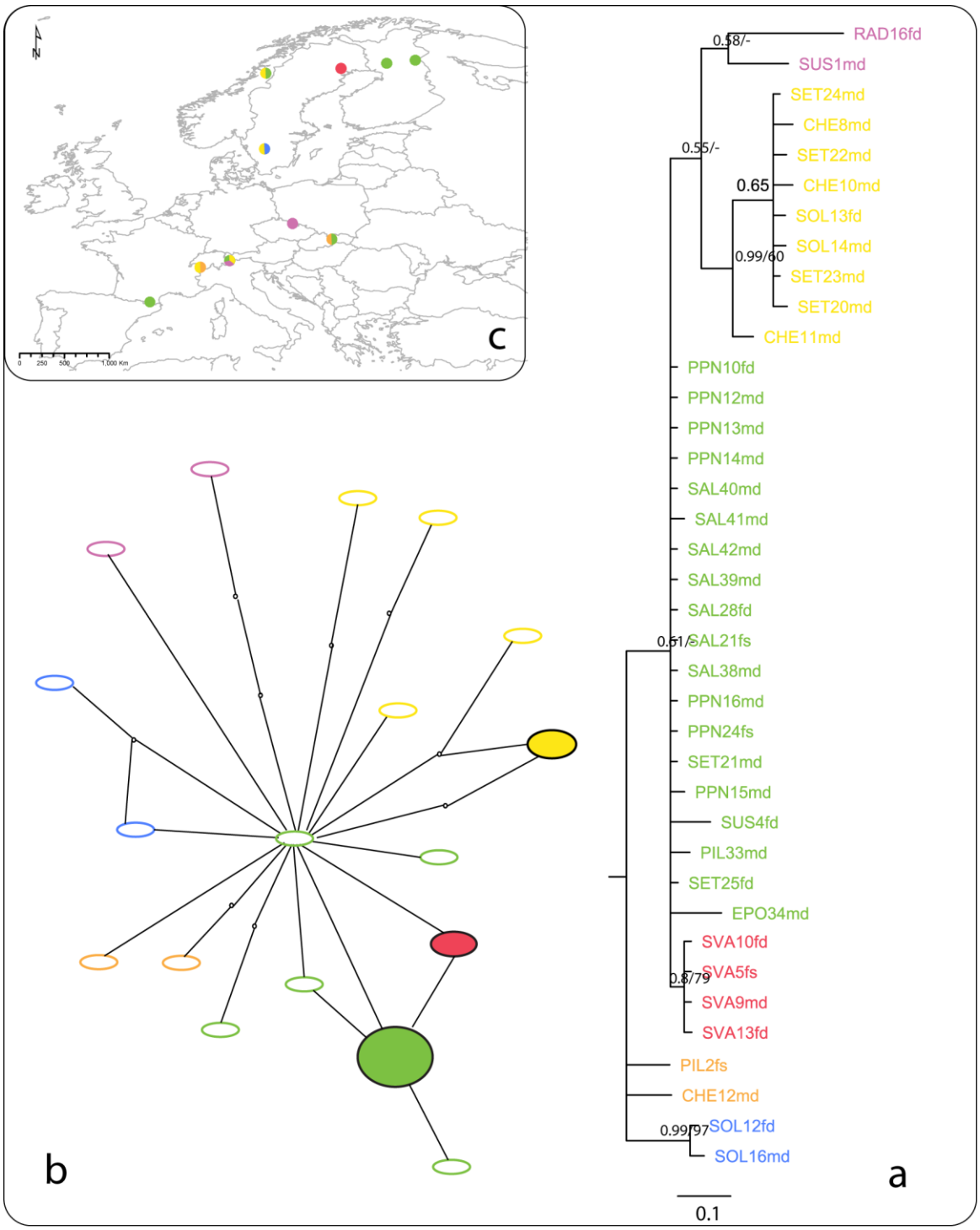
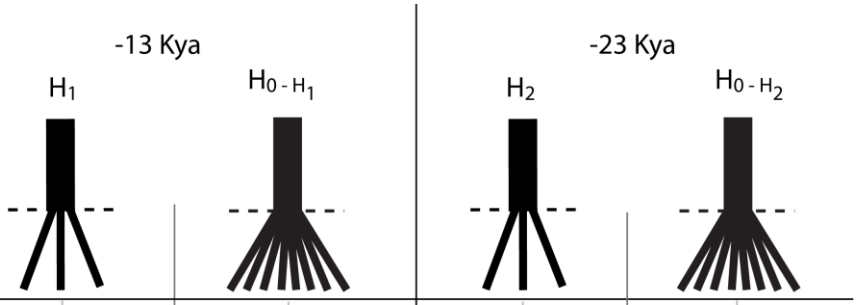


Figure 4 – Molecular inferences (a and b) and geographic distribution of haplotypes (c) in *Chiastocheta dentifera*. a- Bayesian topological inference. Values on branches indicate Bayesian supports and Maximum Likelihood supports. Colors indicate main groups. b- Haplotype network. Colors indicate clades identified in a. Empty circles indicate unique haplotypes; filled circles indicate shared haplotypes. c- Geographic distribution of clades. Colors correspond to a.



	S (emp)	$1*10^{-7}$	$1*10^{-8}$	$1*10^{-7}$	$1*10^{-8}$	$1*10^{-7}$	$1*10^{-8}$	$1*10^{-7}$	$1*10^{-8}$
C. dentifera	24	p=0.76 >0.05	p=0.39 >0.05	<0.001	p=0.002 <0.01	p=0.77 >0.05	p=0.52 >0.05	<0.001	<0.001
C. lophota	27	<0.001	<0.001	p=0.39 >0.05	<0.001	<0.001	<0.001	p=0.55 >0.05	<0.001

Figure 5 – Hypothesis testing for *Chiastocheta dentifera* and *C. lophota*. p-values (p) and levels of significance are indicated for each species, scenario and mutation rate (μ) considered. S(emp) indicates the calculated empirical S value for each dataset.

713 Supplementary Material 1 – Locations, identity and number of sampled flies.

Location	Longitude	Latitude	Altitude (m)	Date	C. dentifera	C. lophota	C. rotundiventris
Ambri	8.702920	46.506800	1000	30.5.08	-	2	2
Amot	8.42346	59.62199	481	18.06.07	-	-	-
Bayasse	6.740670	44.308140	N/A	9.06.07	-	2	-
Beistohlen	8.95473	61.20761	729	19.06.07	-	-	-
Bidjovagge	22.478080	69.297780	613	1.8.08	-	-	2
Bobolice	16.59895	53.94660	110	19.6.08	-	-	2
Col de Bonnecombe	3.11410	44.57557	1335	21.05.07	-	-	1
Braas	15.06817	57.09309	209	15.06.07	-	-	-
Col de la Colombière	6.469722	45.98722	1600	14.06.06	-	-	-
Creux du Van	6.741193	46.93526	N/A	29.06.06	-	2	2
Chasseral	7.021299	47.12569	N/A	29.06.06	-	5	5
Chemin	7.089778	46.08993	N/A	26.05.06	4	1	7
Crans-Montana	7.538896	46.34650	N/A	25.05.06	-	1	1
Cressbrook Dale	-1.740410	53.267240	246	10.6.08	-	-	-
Colt Park	-2.352470	54.193650	380	10.6.08	-	-	1
Donovaly	19.230680	48.889220	1345	21.6.08	-	3	2
Eidda Pastures	-3.741900	53.037200	234	9.6.08	-	-	-
Ellingsrudelva	10.91844	59.91771	161	18.06.07	-	-	-
Esposouille	2.094500	42.623410	1521	3.6.08	1	1	2
Froson	14.60268	63.18205	424	20.06.07	-	-	-
Col du Galibier	6.438611	45.08528	2000	14.06.06	-	4	2
Glen Fender	-3.794850	56.781380	345	14.6.08	-	-	4
Haute Tinee 1	6.818710	44.296170	2020	4.06.07	-	3	-
Haute Tinee 2	6.855810	44.284260	1770	4.06.07	-	-	-
Krasno Polje	14.97271	44.808690	1487	9.7.08	-	-	-
Laktatjakka	18.40674	68.42931	453	22.06.07	-	-	7
Lough Fern	-7.711300	55.065690	26	13.6.08	-	-	-
Loser	13.78485	47.66052	1598	04.06.07	-	5	5
Col long de Magnabaigt	-0.43646	42.87060	1615	23.7.08	-	-	3
Moerlimatt	8.07760	47.90597	939	10.06.07	-	1	1
Monte Pizi	14.167140	41.915240	1546	28.5.08	-	-	-
Naverdal	10.13002	62.70417	480	19.06.07	-	-	-
Pajino Preslo	20.819700	43.277990	1802	8.7.08	-	-	3
Puerto de Panderruedas	-4.972230	43.127430	1227	4.6.08	-	-	4
Pila	20.294490	48.900170	972	21.6.08	2	2	-
Podlesok	20.351900	48.949620	579	21.6.08	-	-	1
petit papa noel	25.79386	66.51647	72	23.06.07	7	-	4
Puy de Dome	2.963333	45.77222	1460	15.06.06	-	2	1
Puy de Sancy	2.809722	45.53500	1520	16.06.06	-	6	1
Radkow	16.353210	50.468660	712	20.6.08	1	-	7
Risnjak - Snjeznik	14.584940	45.438710	1466	5.06.08	-	-	-

Salla	28.65427	66.83020	194	23.06.07	7	-	3
Sede de Pan	-0.486510	43.039490	1556	22.7.08	-	3	1
Seterasen	13.67744	65.53432	285	21.06.07	6	-	1
Solberga	13.56116	57.95194	239	15.06.07	4	-	3
Steingaden	11.01296	47.59529	1158	30.05.07	-	-	1
Straumen	15.64921	67.38440	75	21.06.07	-	-	3
Susch	10.074725	46.747277	N/A	05.06.06	2	2	1
Svartla	21.22062	65.99583	36	22.06.07	4	-	-
Tarasp	10.250556	46.777299	N/A	05.06.06	-	2	-
Trollblumenwiese	11.04118	51.68314	488	16.6.08	-	-	2
Vitosha	23.293420	42.590320	1779	3.7.08	-	-	-
Zali Log	14.110800	46.203420	533	17.5.08	-	-	2

714

General Discussion

Because of the diverse nature of topics, approaches and conclusions addressed in the eight chapters of my thesis, I have chosen to group milestones of my research into the three following themes: i) the species concept in a (co)evolutionary framework; ii) the biology of interactions; iii) the phylogeography of interactions.

Finally, in order to have an insight on the advances and questions opened by this work, I present a small review on the main topics achieved considering each biological system separately. I hope that this approach will not only help resume the work done, but also present some of the new paths that research should take if future work on each system is to be performed.

The species concept in a (co)evolutionary framework

When studying coevolutionary processes it is important to identify which components are particularly driven by selective forces. This aspect has been widely discussed and is intimately related to the species concepts (Wheeler and Meier 2000). Several chapters of this thesis focus on the use of comparative approaches to corroborate and clarify the delimitation of taxonomic groups currently under discussion. Even though involving unrelated groups, the cases of *Arum* (chapter one), *Chiastocheta* (chapter six) and Psychodinae (chapter two) are similar in the approaches used to delimit taxonomic units. In the three chapters, we have demonstrated that the complementary use of different techniques (morphology, molecular phylogeny and spatio-temporal analyses) allows a better understanding and definition of the circumscription and evolutionary history of taxonomic units.

For instance, we have shown that the definition of species in *Arum* is not entirely supported by molecular analyses, indicating that some revision should be undertaken in this genus. Our study has also demonstrated that there are important geographic and historical components in the diversification and relationships among lineages within this genus, a

scenario that was formerly proposed by some authors on the base of distributional evaluations (Hruby 1910).

In the case of species within *Chiastocheta* found to interact with *Trollius europaeus*, limits of morphological species are not always confirmed using molecular markers. Despite the existence of morphological variation between types, some of them frequently hybridize or share considerable ancestral polymorphism. It is thus highly likely that what we currently observe is the result of an ongoing speciation process, in which barriers to hybridization are still not completely defined. As mentioned by de Queiroz (2005), our study thus points out the fact that the main problem in the discussion of the species concept is related to the fact that different techniques evaluate entities formed at different timescales. In this study, we demonstrate that depending of the criterion applied (morphological, mating isolation, molecular-based phylogenetic monophyly), different supported entities are identified. Our study provides a first large-scale survey of the relationships between lineages within this fly group, considering not only their genetic identity, but also their distribution and possible gene flow among species.

Finally, the case of Psychodinae occurs at a higher taxonomic level. Here, we demonstrated that despite their relatively similar morphologies, the subfamily Psychodinae is extremely old (up to 85 My). We also showed that only some aspects of their current systematics (*i.e.,* some subtribal entities, some genera) are confirmed by molecular analyses. Moreover, it appears that in this group most species lineages originated in the Paleogene, shortly after the K-T massive extinction event (Erwin 2001; Jablonski 2001).

These three case studies have in common the fact that taxonomic circumscriptions of entities treated in them were problematic. In all of them, the diverse approaches applied helped open up the possibility to discuss concepts in a different framework than is generally employed (e.g. morphological). The more comprehensive approach we used presents thus new options to explain these complex relationships.

In the framework of the species concept discussion, I hope that the work done here will at least partially help provide arguments to demonstrate that the debate over correctness of one concept over another is not necessarily constructive. Each concept has problems and

advantages, and probably the biggest potential lies in exploring and combining them thoughtfully, using each in ways appropriate to the organisms under study. A good demonstration of this may be the fact that within these chapters, any approach by itself (*e.g.*, molecular, morphological, biogeographic) was able to independently provide convincing propositions in regard to the systematics of the discussed groups, whereas a combination of these approaches brought new light to the boundaries of taxonomic units. Moreover, one of the most exciting points about complementary analyses is that they allow us to position ourselves at different evolutionary and temporal standpoints, to have a deeper understanding of the speciation process.

Ecology and biology of interactions

Coevolutionary processes are characterized by their dynamic nature: species are continuously adapting to each other in order to increase their respective fitness gains (Bronstein, Dieckmann et al. 2004). Despite the fact that coevolutionary processes involve a defined number of species (Ollerton, Killick et al. 2007), organisms participating to the interaction are influenced both by their abiotic environment and by pressures exerted by the other interacting organisms. Investigating the effects of abiotic and biotic aspects on the dynamics of biological and ecological adaptations was the aim of chapters three and four.

First, the Theory of the Geographic Mosaic of Coevolution (TGMC) predicts that in a variable environment, biotic interactions are susceptible to change (Thompson 2005). Taking this idea into account, I demonstrated in chapter three that there is an effect of climatic variation on the identity of potential pollinators in the antagonistic interaction represented by *A. maculatum* and its Psychodid pollinators. As predicted by the theory, it appears that the plant switches pollinators according to clinal changes in precipitation and temperature. Besides confirming the TGMC, this result demonstrates the importance of large-scale studies when working with specific interactions and warns against assuming general laws for processes which occur locally and are punctually confirmed.

In the same way that the environment can influence the nature of a given interaction at a large spatial scale, theory predicts that interactions can force associated species to maximize their benefits from the ecological exchange (Dufaÿ and Anstett 2003). For instance, plants maximize fitness by adapting their reproductive strategy based on the type and efficiency of their pollinators. In chapter four, we showed that the identity and diversity of pollinators appears to shape the reproductive strategy in a series of closely related plants of the genus *Arum*, which all present entomophilous lure-and-trap pollination systems. In this chapter, we test a simple model to help clarify our understanding of the relationship of the reproductive strategy represented by the Size-Advantage Model (Ghiselin 1969) and the type of pollinators a plant is associated.

Probably the most important demonstration of these chapters is that coevolutionary interactions should not be seen as static, even if they appear to be extremely specialized. Variations in the environment or in the respective adaptation of partners can modify selective forces and thus redirect selective pressures, either locally or along the whole geographical range of an organism.

Phylogeography of interactions: understanding interactions in time and space

The field of phylogeography is an integrative field *per se*, since it combines molecular phylogenetic inferences with spatial information (e.g., geographical information systems, spatio-temporal distribution of species; Hickerson, Carstens et al. 2010).

In this context, the comparative phylogeography of interactions goes further by integrating hypotheses that explain phylogeographic patterns of species interacting ecologically and thus integrating ecological factors into the analysis of the evolutionary history of species.

In the same way that understanding the history of species is important for recognizing the effect of external events (e.g., climatic fluctuations, orogenic processes) on their survival (Hewitt 1996), comparing phylogeographic patterns in ecologically related species informs us about the influence that the interaction itself had on the history of species (Tsai and

Manos in press). Knowing the effect one species could have on another thus allows the quantification of the interdependency of species in a temporal framework.

In such a context, in chapters five, seven and eight, I studied the comparative phylogeographies of species involved in two European, obligate and specific interactions. Considering the specific and obligate nature of the relationships, I initially expected that while mutualistic interactions would be prone to share similar (congruent) phylogeographic histories, this would not be true for antagonistic interactions.

Our results partially confirmed these working hypotheses, since antagonistic organisms presented incongruent phylogeographic patterns (chapter five), while some of the interacting species involved in the mutualistic interaction presented congruent phylogeographies (chapter eight).

In chapter five, the finding of incongruence between antagonistic species could be related to differences in generation times and dispersion capabilities between the two partners [as has been proposed by Borer *et al.* (submitted)]. It is interesting to note that the phylogeographic structure demonstrated by the plant is extremely similar to that presented by the composition of flower visitors (chapter three). If the current insect composition has been brought about by the last glacial period, it is possible that the plant genetic structure has been shaped by adaptation to different fly species, which would have been hosted in different glacial refugia.

In chapter eight, our results on the mutualistic interaction between *T. europaeus* and its *Chiastocheta* pollinators showed that while species are similar from an ecological point of view, their evolutionary histories appear to be different. This could be a consequence of interspecific competition between species exploiting similar resources. This struggle could have induced different taxa to develop diverse dispersal capabilities and developmental features to displace their respective ecological niches and avoid competition. Our results thus demonstrated that while the genetic pattern of one species (*i.e.*, *C. lophota*) appeared to be strongly related to geographic features corresponding to phylogeographic patterns already identified for other Alpine species (e.g., Schönswetter, Stehlik et al. 2005), the other

two presented almost no spatially defined genetic structure (*C. rotundiventris*), or had a structure which appears to has been driven by the host plant (*C. dentifera*).

While a wide range of causes appeared to explain the phylogeographic patterns of *Chiastocheta* flies, the plant's spatial genetic structure is more likely explained by its distribution at the Last Glacial Maximum (LGM): there is hence a strong effect of the abiotic component in the definition of the plant structure. In the future, it would be interesting to test the plant structure using phylogeographic statistical techniques similar to those used for the insects. This is, however, currently technically impossible considering the type of genetic data (dominant) we used.

These three chapters demonstrate that complex phylogeographic questions, such as those involving congruence testing in the context of interacting species, might require going further than just visually quantifying similarities between distribution maps of genetic diversities. Today, an array of new techniques and technologies is available to biologists, and the potential for understanding the history of evolutionary processes in space is enormous. In this thesis for example, the integration of several areas of expertise has allowed the development of a new technique to infer past demographic contractions in cold-adapted species (chapter seven), and the direct testing of phylogeographic scenarios applied in chapter eight. Without such interdisciplinary work, none of the conclusions obtained in those surveys would have been possible.

State of the knowledge on each biological system

Arum maculatum L. – Psychodid flies

Achievements attained in this thesis allowed a new understanding of evolutionary and biological topics of this antagonistic pollination relationship.

From the plant evolutionary point of view, we could demonstrate that the taxon previously considered as *A. maculatum* should be redefined in the future. Our study (chapter one) indicated that not one but two phylogenetic entities are currently being considered as belonging to this species. While one of them appears restricted to the European part of the

distribution of the plant, individuals from Anatolia seem to belong to another taxon. It appears necessary that these two entities are further investigated to clearly define their taxonomic boundaries. Moreover, whereas the Anatolian entity appears as monophyletic, this is not the case of the European taxon, whose specimens are included into a wide polytomy comprising other *Arum* such as *A. cylindraceum*. In the future, investigating species definition using more variable markers should allow addressing the monophyly of *A. maculatum*.

From the biological point of view, we demonstrated that the number and identities of pollinating flies appear to be geographically structured throughout the European part of the distribution range of the plant (chapter three). While *Psychoda phalaenoides* L. is the dominant visitor in central and northern Europe, populations in southeastern regions are dominated by *Psycha grisescens* Tonnoir. This result is of main importance because it demonstrates that even in highly specialized biological systems as the one studied here, variation in terms of abundances and identities of visitors is also expected when the distributions of interacting species cover large spatial ranges. This study also points out the importance of having a clear view on the natural history of the system studied before further investigating other more complex evolutionary questions.

The main advances done at the phylogeographic level are that even though the plant has a clear geographic genetic structure, this is not the case of the pollinators (chapter five). Despite this incongruence at the level of intraspecific lineages, the spatial structuring of the pollinator compositions is similar to the plant genetic structure, what would indicate some correlation between these two components. It is possible that differences in insect compositions are currently shaping the genetic structure of the plant through differential gene flow defined by the two pollinator groups. Another option could be that plants have been adapting to the two types of visiting insects through glacial times and that what we currently see is the result of selective and adaptive processes. Because plants attract insects by odour and heat production, it would be interesting in the future to investigate the volatile compositions of plants belonging to different regions. If adaptive processes were

happening, we would expect that plants from the two regions would produce different volatiles, more adapted to the attraction of the corresponding insect species.

At the same time, and in order to be sure that the flower chambers compositions do not reflect only the species availabilities in the environment, general insect trapping surveys should be done at the local level. We could thus understand with more accuracy the variation and compositions of insect availability in the two different environments that these plants occupy.

***Trollius europaeus* L. – *Chiastocheta* species complex**

Probably one of the main advances on this mutualism was that thanks to our large-scale sampling strategy, we could provide a sight onto the insect visitation rates throughout a large part of the distribution range of the plant (chapter six). This approach allowed us to observe that while some species currently occupy the whole European distribution of the plant, this is not the case of all of them. However, by using molecular techniques to identify lineages and congruence between molecular and morphological approaches, we showed that even though some species are clearly genetically circumscribed, this is not the case of all of them, with especially four species presenting relatively high levels of genetic admixture. Besides demonstrating that the species boundaries in these closely related set of species are difficult to define, this study warns researchers on conclusions that could be obtained if the species are considered to be clearly isolated entities. From this point of view, it should be important in the future to take these results into account, especially if working on the evolutionary and ecological dynamics of those admixed entities.

From a phylogeographic point of view, we demonstrated that the genetic structure of the plant appears to have been affected by the climatic variations related to the end of the last glacial period (chapter seven). From the fly side, only one species (*C. dentifera*) presented a phylogeographic structure similar to that of the plant (chapter eight), what could indicate that this species is more dependent of the interaction than the others, probably because it is the one visiting plant patches at the end of the flowering period.

The next steps in the understanding of this system would need a deeper survey on the detailed life-cycles and dispersion capabilities of the different interacting species. It would be moreover also interesting to investigate the interspecific gene flow among the different fly morphospecies.

From the fly community point of view, it would be important to study the behavior of those species present at locations in which most of the other taxa are lacking. It could be expected that because of differences in fly interspecific competition pressures at each location, taxa might behave differently when their relative diversities vary.

Finally, from a taxonomical point of view, it appears necessary to redefine the boundaries of the two species appearing to present the highest rates of interbreeding (*C. setifera* and *C. inermella*). For this, collaborations with taxonomists are highly recommended, since their knowledge could largely help find informative morphological characters.

References

Borer, M., S. Buerki, et al. (submitted). "Comparative phylogeography in a specific plant-insect antagonism: partial congruence in large-scale spatial genetic structures associated with similar dispersal abilities." Molecular Ecology.

Bronstein, J. L., U. Dieckmann, et al. (2004). Coevolutionary dynamics and the conservation of mutualisms. I. I. f. A. S. Analysis. Laxenburg, International Institute for Applied Systems Analysis**:** 1-23.

de Queiroz, K. (2005). "Ernst Mayr and the modern concept of species." Proceedings of the National Academy of Sciences of the United States of America **102**: 6600-6607.

Dufaÿ, M. and M.-C. Anstett (2003). "Conflicts between plants and pollinators that reproduce within inflorescences: evolutionary variations on a theme." OIKOS **100**: 3-14.

Erwin, D. H. (2001). "Lessons from the past: biotic recoveries from mass extinctions." Proceedings of the National Academy of Sciences of the United States of America **98**(10): 5399-5403.

Ghiselin, M. T. (1969). "The evolution of hermaphroditism among animals." Quarterly Review of Biology **44**(2): 189-208.

Hewitt, G. M. (1996). "Some genetic consequences of ice ages, and their role in divergence and speciation." Biological Journal of the Linnean Society **58**(3): 247-276.

Hickerson, M. J., B. C. Carstens, et al. (2010). "Phylogeography's past, present, and future: 10 years after Avise, 2000." Molecular Phylogenetics and Evolution **54**(1): 291-301.

Hruby, J. (1910). "Le genre *Arum* - Aperçu systématique avec considérations spéciales sur les relations phylogénétiques des formes." Bulletin de la Société Botanique de Genève **4**: 330-371.

Jablonski, D. (2001). "Lessons from the past: evolutionary impacts of mass extinctions." Proceedings of the National Academy of Sciences of the United States of America **98**(10): 5393-5398.

Ollerton, J., A. Killick, et al. (2007). "Multiple meanings and modes: on the many ways to be a generalist flower." Taxon **56**(3): 1-12.

Schönswetter, P., I. Stehlik, et al. (2005). "Molecular evidence for glacial refugia of mountain plants in the European Alps." Molecular Ecology **14**(11): 3547-3555.

Thompson, J. N. (2005). "Coevolution: the geographic mosaic of coevolutionary arms races." Current Biology **15**(24): 992-994.

Tsai, Y.-H. E. and P. S. Manos (in press). "Host density drives the postglacial migration of the tree parasite, *Epifagus virginiana*." Proceedings of the National Academy of Sciences of the United States of America.

Wheeler, Q. D. and R. Meier, Eds. (2000). Species concepts and phylogenetic theory. New York, Columbia University Press.

General Conclusion

The development of different approaches in the biological sciences has made this area extremely dynamic. The diversification of studies has dramatically increased the complexity of topics investigated, but has opened exciting new doors towards the establishment of interdisciplinary studies.

This thesis has exploited such integration of approaches, by basing hypotheses and analyses on different ecological, evolutionary, phylogeographic and historical aspects of interactions, while trying to incorporate the knowledge accumulated in different areas of study in biology. As demonstrated in this work, this is useful not only for the development of new analytical approaches, but also for cases of taxonomic delimitations, especially when incongruence between different types of data is present.

Thanks to such approaches, we are able to answer to the main question of this thesis: do interacting partners present similar histories? The answers were yes for some mutualistic and no for antagonistic relationships. However, in both cases, we demonstrated that relationships between interacting species are not as simple as previously thought, and that the simultaneous and complementary analysis of different aspects can help enhance our view of interactions.

I am convinced that the future of biological research relies on collaboration, exchange and communication between researchers working on different topics. In the same way that this thesis has benefited of interdisciplinary work and scientific flow among researchers and institutions, numerous are the areas of evolution and ecology which should take advantage of this to investigate and understand the complexity of life.

Appendix 1

Curriculum vitae

CURRICULUM VITAE

Espíndola, María Anahí
Chemin des Liserons 9
2000 Neuchâtel
Switzerland
Phone: +41 (0)32 724 4779
maria.espindola@unine.ch

Biologist
28 years-old
married
Argentine

Studies

2006 - 2010	PhD thesis at the E-vol Laboratory of the University of Neuchâtel, Switzerland: "Inferring reciprocal evolutionary histories in associated species of plants and insects in two European pollination systems", under the direction of Prof. Martine Rahier and Dr. Nadir Alvarez.
2002 - 2006	Biology, diploma at the E-vol Laboratory of the University of Neuchâtel, Switzerland. Diploma: "Interactions between *Horismenus butcheri* Hansson & Aebi (Hymenoptera: Eulophidae) and two *Phaseolus* (Fabaceae) bean species in Central Mexico. Biology, population genetics and geospatial studies", under the direction of Dr. Betty Benrey.
2000 - 2001	Biology, at the National University of Córdoba, Argentina.

Courses

January-March 2010	"Computational Phylogenetics and Phylogeography". Prof. Bryan C. Carstens. Louisiana State University, USA.
13-28 May, 2009	"An introduction to the Practice of Statistics using R". Dr. Thomas Gsponer. NCCR Doctoral Program. University of Neuchâtel, Switzerland.
10-14 September, 2007	"Phylogénie et évolution moléculaire". Dr. Juan Montoya and Dr. Jan Pawlowski. Departement of Zoology and Animal Biology. University of Geneva, Switzerland.
9-11 October, 2006	"Population genetic data analysis". Dr. Felix Gugerli and Dr. Rolf Holderegger. Plant Science Center Zürich. ETHZ, Switzerland.

Conferences

13-16 October 2009	Poster. DIVERSITAS Open Science Conference 2. "Biodiversity and Society: Understanding connections, adapting to change", Cape Town, South-Africa.
8-10 July, 2009	Oral communication. “Biogeography of *Arum maculatum* pollinators”. Xth International Aroid Society Conference . Nancy, France.
8-10 July, 2009	Poster. “Pollination biology of *Arum cylindraceum*”. Xth International Aroid Society Conference . Nancy, France.
January 30 – February 1, 2008	Participation at the “Plant Species Concepts and Evolution”: NCCR Plant Survival Conference. University of Neuchâtel, Neuchâtel, Switzerland.
17-21 September, 2007	Poster. “Population and landscape genetics of a Mexican parasitoid”. Xth European Workshop on Insect Parasitoids, Erice, Italy.

Publications

Espíndola A., Buerki S., Bedalov M., Küpfer P. and Alvarez N. 2010. New insights into the phylogenetics and biogeography of *Arum* (Araceae): unravelling its evolutionary history. *Botanical Journal of the Linnean Society*. 163, 14–32.

Espíndola A., Pellissier L. and Alvarez N. *in press*. Variation in the proportion of flower visitors of *Arum maculatum* along its distributional range in relation with community-based climatic niche analyses. OIKOS. doi: 10.1111/j.1600-0706.2010.18937.x

Revel N., Alvarez N., Gibernau M. and **Espíndola, A.** *In preparation*. Size-advantage model and pollination strategies in plants. To be resubmitted to Evolutionary Ecology.

Espíndola A., Pellissier L. and Alvarez N. *In preparation*. No fossils, no answers? Inferring postglacial genetic consequences in cold-adapted species when neither ancient DNA nor macroremains are available.

Espíndola A. and Alvarez N. *In preparation*. Comparative phylogeography in a specific and obligate pollination antagonism.

Espíndola A., Buerki S., Jacquier A., Ježek J. and Alvarez N. *In preparation*. Molecular relationships in the sub-family Psychodinae (Diptera: Psychodidae). To be submitted to Molecular Phylogeny and Evolution.

Espíndola A., Buerki S. and Alvarez N. *In preparation.* Evolutionary history of *Chiastocheta* species interacting with *Trollius europaeus*.

Espíndola A., Carstens B. C. and Alvarez N. *In preparation.* Phylogeography of mutualists and the effect of the host on the genetic structure of its partners.

Espíndola A., Revel N. and Alvarez N. *In preparation*. Phylogeography of *Arum cylindraceum*.

Triponez Y., **Espíndola A.**, Bassin L., Pellissier L. Alvarez N. *In preparation*. Large biogeography of an oil collecting mutualism: tight interactions involve shared refugia but independent dispersal of partner species.

Internships

15.08.05-09.09.05	Study of gene flow between different species of the genus *Aegilops* in Europe, using RAPD markers: "*La génétique de populations d'Aegilops neglecta Req. et Bertol., A. cylindrica Host, A. triuncialis L. et A. geniculata Roth dans quatre pays d'Europe*". Laboratoire de Botanique Evolutive. University of Neuchâtel. Under the direction of Dr. Roberto Guadagnuolo.
26.07.04-10.09.04	Primer creation and sequencing of genes involved on species isolation in *Drosophila montana:* "Mating song evolution of *Drosophila montana* (Diptera: Drosophilidae). A molecular analysis of the evolution of the cacophony (*cac*) gene*".* Laboratory of Biology and Evolution, the University of Leeds (UK). Under the direction of Prof. R. K. Butlin.

Teaching

2007 - 2009	Bachelor classes at the University of Neuchâtel: "Ecologie Evolutive", "Méthodes Quantitatives en Ecologie" and "Biologie des Invertebrés". Master classes at the University of Neuchâtel, Master of Physiology and Ecology of Plants (PEP): "Classics in Biology".

Field experience

April – August, 2007 April – August, 2008	Field trips covering the European continent (Great Britain, Scandinavia, Balkans, Central Europe, Italy and Spain).
26.01.2005-26.02.2005 31.12.2005-8.02.2006	Field trips in Central Mexico.

Prizes

October, 2006	Jean Landry Prize, for a mean of 5.73 at the end of the studies.

Grants

2009-2010	*Fonds des Donations* of the University of Neuchâtel, to get training on methods in phylogeographical statistics at the Laboratory of Bryan C. Carstens, Department of Biological Sciences, Louisiana State University, USA.
April – June 2008	SCNAT+ funds for field work, to carry out the sampling of *Arum cylindraceum* and its pollinators throughout its South-East European distribution
July – September 2004	Funds Wüthrich - Matthey-Dupraz; for an internship at the Laboratory of Prof. R. K. Butlin, Leeds, UK.

Informatics

- → Wide experience with Geographic Information Systems: ArcGIS and ArcView
- → Good knowledge in exploration of Biodiversity and geographic databases: GBif, EUNIS, several European local databases, European soils database.
- → Good experience in the use of population genetics and phylogenetics softwares: e.g., STRUCTURE, Geneland, MrBayes, F-stats, Arlequin, TreeFinder, MrAIC, PAUP*, Mesquite.
- → Good knowledge in the use of graphic treatment of images (Photoshop and Illustrator).

Languages

- → Spanish: native speaker.
- → English: very good written and spoken knowledge (Upper Intermediate Level).
- → French: excellent written and spoken knowledge ("Accès au DALF" certificate + 8 years of studies at the University of Neuchâtel, Switzerland)
- → German: good spoken (Swiss-German) and some written knowledge.