Evolutionary history of aphid-plant associations and their role in aphid diversification

Histoire évolutive des associations plantes-pucerons et leur rôle dans la diversification des pucerons

Jean Peccoud a,*, Jean-Christophe Simon b, Carol von Dohlen c, Armelle Coeur d'acier d, Manuel Plante genest b, Flavie Vanlerberghe-Masutti d, Emmanuelle Jousselin d

a INRA, UMR BGPI, Campus international de Baillarguet, TA A-54/K, 34398 Montpellier cedex 5, France
b INRA, UMR BiO3P, agrocampus Ouest, université Rennes 1, domaine de la Motte, 35653 le Rheu cedex, France
c Department of Biology, Utah State University, 5305 Old Main Hill, Logan, Utah 84322, USA
d INRA, UMR CBGP, IRD, Cirad, Montpellier SupAgro, Campus international de Baillarguet, CS 30016, 34988 Montferrier-sur-Lez cedex, France

ARTICLE INFO

Article history:
Available online 13 May 2010

Keywords:
Aphididae
Speciation
Host-plant association
Ecological speciation
Phylogenetic reconstruction

ABSTRACT

Aphids are intimately linked with their host plants that constitute their only food resource and habitat, and thus impose considerable selective pressure on their evolution. It is therefore commonly assumed that host plants have greatly influenced the diversification of aphids. Here, we review what is known about the role of host plant association on aphid speciation by examining both macroevolutionary and population-level studies. Phylogenetic studies conducted at different taxonomic levels show that, as in many phytophagous insect groups, the radiation of angiosperms has probably favoured the major Tertiary diversification of aphids. These studies also highlight many aphid lineages constrained to sets of related host plants, suggesting strong evolutionary commitment in aphids' host plant choice, but they fail to document cospeciation events between aphid and host lineages. Instead, phylogenies of several aphid genera reveal that divergence events are often accompanied by host shifts, and suggest, without constituting a formal demonstration, that aphid speciation could be a consequence of adaptation to new hosts. Experimental and field studies below the species level support reproductive isolation between host races as partly due to divergent selection by their host plants. Selected traits are mainly feeding performances and life cycle adaptations to plant phenology. Combined with behavioural preference for favourable host species, these divergent adaptations can induce pre- and post-zygotic barriers between host-specialized aphid populations. However, the hypothesis of host-driven speciation is seldom tested formally and must be weighed against overlooked explanations involving geographic isolation and non-ecological reproductive barriers in the process of speciation.

© 2010 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.
1. Introduction

Most insect herbivores have very intimate associations with their host plants. Debates continue over the extent to which the diversification of phytophagous insects is driven by their association with plants [1]. Several hypothesized pathways of diversification recognize host plants as playing a major role in the process of speciation; two of these are:

(i) cospeciation–herbivores diversify contemporaneously with their host plants, resulting in parallel phylogenetic trees between insects and their hosts [2,3];
(ii) speciation through host shifts—a group of individuals of an insect species colonizes a new existing plant species and diverges from its population of origin [4].

Both processes rely on some degree of host plant specialization and suggest that host plant diversity, if not necessarily the sole cause of reproductive isolation and speciation, is a major driver of insect diversification [5].

There are approximately 4700 described aphid species (Aphididae, sensu Remaudière and Remaudière [6]). Forty percent of aphid species live wholly or partly on trees and 55% live on herbaceous plants or shrubs (the remaining 5% live on unknown hosts) [7]. Although the host breadth of aphids is variable, aphids are recognized as being intimately linked with their host plants [8]. This statement stems from several well-documented observations. First, aphids and their host plants are not haphazardly associated. Some plant families are far overrepresented among aphid hosts, such as Asteraceae (> 600 aphid species), Apiaceae (~600), Rosaceae (~300) and Coniferae (~350) [7]. Conversely, some large plant families (e.g. Cactaceae, Euphorbiaceae, Orchidaceae, Lauraceae, Moraceae, Myrtaceae, [9]) host very few aphid species. Second, among all phytophagous insects, aphids are generally recognized as being near the end of the continuum from generalists to specialists. Specialization occurs at various taxonomic levels. Some aphid species are specific to a single plant species. At a higher taxonomic level, some aphid genera/families are strictly associated with a single plant genus/family. However, several highly polyphagous aphid species are also known (less than 1% of all species), including some of the most injurious agricultural pests. For instance, Aphis gossypii Glover can feed on 912 species plants belonging to 116 families [10]. However, apparently polyphagous species frequently encompass host races or host-specialized populations with a narrow host range [11–15]. The detection of several differentiated host races within apparently polyphagous species suggest an ongoing process of speciation by adaptation to distinct host plants in several aphid groups.

Altogether, these patterns of host use prelude the idea that host-plant associations have influenced aphid diversification. Here, we evaluate this hypothesis by documenting the history of host-plant association and specialization through phylogenetic reconstructions, and by reporting on how host plants can promote reproductive isolation in natural populations.

2. Macroevolutionary patterns of aphid-plant associations

2.1. Major evolutionary transitions in the history of aphid-plant associations

Aphididae are divided into several subfamilies, tribes and subtribes (we use here the most recent comprehensive classification of [6]). Among subfamilies and tribes, few are strictly associated with gymnosperms, some use herbaceous monocotyledons and most use various angiosperm
families. Few molecular phylogenies have investigated the origin of these tribes and their relationships; we have attempted here to summarize these studies and what they tell us about the evolution of host plant association (Fig. 1).

Phylogenetic reconstruction places the coniferous feeding family of Adelgidae as a sister group of Aphididae [16]. Several lineages of Aphididae also use coniferous hosts, and these may represent ancient associations [17]; one such lineage is the subfamily Neophyllaphidinae, which has a Gondwanan distribution and is associated with two gymnosperm genera, Podocarpus and Araucaria. It has been suggested that early Aphididae ancestors lived on coniferous hosts, a hypothesis consistent with fossil records [18,19]. For example, Neophyllaphidinae look very similar to the Cretaceous fossil Aniferella, which by inference were probably associated with Podocarpaceae or Araucariaceae [20]. These two gymnosperm families were likely ancestral aphid hosts in the Early Cretaceous before angiosperms arose, along with several other abundant gymnosperm groups [21]. Fossils have allowed the calibration of molecular phylogenetic reconstructions: the age of the common ancestor of extant subfamilies in the Aphididae has been estimated at 84–99 mya [21], which corresponds to the Late Cretaceous epoch and roughly the diversification of woody angiosperms. The lack of resolution of molecular phylogenetic relationships among Aphididae subfamilies and tribes and the short branches found in the deep nodes of the phylogenetic tree suggest a rapid radiation of aphids at this period [21,22]. A more recent study found tentative support for three main lineages, with Lachninae as sister group to all other aphids; as one of the few subfamilies with conifer feeders, the position of Lachninae also supports the hypothesis of an ancestral gymnosperm feeding in aphids [23] with the caveat that conifer feeding in lachnines is thought to be a derived state by some systematists [24,25]. That most of present-day aphid diversity likely arose after one or more shifts from gymnosperms to angiosperms strongly suggests that the radiation of aphids has been driven by the

---

**Fig. 1.** Putative phylogenetic relationships of Aphididae sub-families, adapted from von Dohlen and Moran [21] using the classification of Remaudière and Remaudière [6] and results from references [22,23]. * indicates groups in which monophyly can be contested based on molecular phylogenies. References are indicated for subfamilies for which molecular phylogenies have been conducted: 1[53], 2[52], 3[51], 4[141], 5[21].
diversification of flowering plants [26]. Fewer than 30 extant species remain in two subfamilies thought to be primitively conifer feeding (Neophyllaphidinae and Mindarinidae), compared to more than 400 angiosperm-feeding species. These observations suggest that either aphids could have speciated in parallel with particular groups of angiosperms (i.e., have partly cospeciated) [27] or that their diversification has rapidly followed the radiation of angiosperm families, favoured by the diversity of new ecological niches offered by these plants.

Compared to other phytophagous insect groups of similar age, aphid biodiversity is rather low. Furthermore, aphids are relatively rare in the tropics, where plant diversity is the highest, and underrepresented in the southern hemisphere. The plant families cited by Eastop [24] and Dixon [9] as having no specific aphids are mostly diversified in tropical or hot regions and/or in the southern hemisphere. Thus, absence of diversification on certain plant families may have little to do with aphid feeding preferences. Several hypotheses might explain the paucity of aphids in the tropics and southern hemisphere. One of them suggests that aphids have been less successful in tropical regions due to the difficulty of finding their specific host plants in the highly diversified tropical ecosystems; aphids seem to locate their hosts somewhat randomly and may depend on high plant apparency to do so [9,28]. Another hypothesis suggests that cyclical parthenogenesis, the ancestral life cycle of all sexually reproducing species, is maladapted to the tropics [29]. Low diversity in the southern hemisphere compared to the northern one has also been explained in a historical biogeographic context, predicated on the idea that aphids originated in the northern hemisphere and underwent extensive species radiations there during the Tertiary Period; subsequently the tropics acted as a barrier to southward migrations [30]. Aphid geographical distribution and plant association is most certainly the result of both ecological and historical constraints. Comparative biogeographical studies of aphids and their host plants integrating advances in molecular phylogenetics are still quite rare [31,32], and more are needed to elucidate several of the major transitions in aphid diversification.

Host alternation, a characteristic trait of certain aphid life cycles, may have played an important role in favouring the colonization of new plants and speciation in aphids. About 10% of aphid species alternates seasonally, and mostly obligately, between two unrelated plant species [7,33,34]; these aphids are said to be host alternating or heteroecious, in contrast to the vast majority of species that complete their entire life cycle on the same host (known as non-host-alternating or monoeocious species). The evolution of host alternation has occurred in several aphid families, in which the cycle follows a roughly similar pattern of host use: after egg hatching followed by parthenogenetic development of colonies on their primary/winter host (most generally a woody host), aphids alternate to secondary/summer hosts (most generally herbaceous plants) on which they reproduce parthenogenetically; later generations return to their primary host for sexual production of eggs and usually overwintering. The two host plants of host alternating species are almost never closely related, and some consist of rather surprising combinations. For instance, the Nearctic species *Melaphis rhois* (Fordini) and the related eastern-Asian species of *Nurudea*, *Kaburagia* and *Schlechtendalia* (Fordini) alternate between *Rhus* (Anacardiaceae) and mosses [35]. The exact number of evolutionary gains and/or losses of this life cycle throughout aphid history is not known, but almost certainly both gains and losses have occurred multiple times, and some proportion of extant aphid-host associations probably have been shaped by such transitions [35]. The evolution of host alternation has two intricately linked consequences:

(i) the colonization of new ecological niches via a new life cycle;
(ii) the broadening of host plant range.

Host-alternating species often have a much larger range of secondary hosts than primary hosts (see section 2.2 for possible causes). Hence, host alternation seems to allow for an escape from host plant specialization. This transition in aphid diet might be followed by a shift of all or part of the population to permanent existence on some of the new secondary hosts, and eventually leading to speciation. These sequences of evolutionary changes could have fuelled the diversification of certain aphid lineages [34].

Such a scenario entailing acquisition of secondary hosts and broadening of host range, followed by specialization and speciation on newly acquired hosts, might be exemplified by the diversification of Aphidinae. With about 57% of all described aphid species [24], Aphidinae constitutes the most species-rich subfamily. It encompasses many genera that live exclusively or partially on Rosaceae and are also associated with modern grasses. It also includes many highly polyphagous species among which most of the agricultural aphid pests are found. The scarcity of Eocene and Oligocene fossils suggests that this lineage probably emerged and diversified in the latter half of the Tertiary (Miocene and later), when plant communities dominated by grasses became abundant [19,26]. Molecular phylogenetic reconstruction implies several evolutionary transitions in life cycles in this subfamily [36]. Further, phylogenies suggest that several shifts to and from Rosaceae have occurred. One plausible evolutionary scenario is that ancestors of Aphidinae fed on a woody host, possibly an early Rosaceae species [27]. This lineage then radiated following the acquisition of host alternation to herbaceous hosts; in some cases descendent lineages lost their primary rosaceous hosts to become monoeocious on more recently acquired secondary hosts (also called secondary monoecy) [8,19,36,37]. Several genera in Aphidinae exhibit all the life cycle states in this hypothetical process (primary monoecy, host alternation, and secondary monoecy), including *Dysaphis*, *Cryptomyzus*, *Myzus*, and *Brachycyclus*. In contrast, no other aphid lineages with host alternation (Eriosomatinae, Hormaphidinae, Anocelinae) exhibit complete secondary monoecy including a sexual generation. Given the disparity in species diversity between Aphidinae and other lineages, it seems possible that acquisition followed by loss of
heterocyclic may have accelerated the diversification process in this lineage [34].

A true comparative analysis of the diversification rate is difficult to conduct based on the current phylogenetic reconstructions, but it would be interesting to test whether increased diversification rates are correlated with acquisition and loss of host alternation. Such comparative tests could be conducted at several taxonomic levels: among subfamilies, and among genera within Aphidinae. These analyses could yield insights into the historical role of evolutionary transitions in life cycle and host plant range in aphid speciation.

2.2. Patterns of host plant specialization: are aphids actually good botanists and why?

As mentioned above, aphids are specialized at several taxonomic levels. More than half of all aphid species feed on a single plant species, and many aphid genera are restricted to a single plant genus or several closely related plant genera [24]. Among the 445 non-host-alternating European Aphidinae, 407 (91%) feed on host plant species confined to a single genus [24]. Such host specificity extends across all aphid lineages: for example, all primary hosts of Pemphigus (Eriosomatinae) belong to Populus [8]; all species of Chaitophorus (Chaitophorinae) are confined to the genera Salix and Populus (both in Salicaceae). Some tribes are restricted to one or more genera within a single plant family; all primary hosts of Eriosomatini belong to two genera of Ulmaceae; Hormaphidini use primary hosts in Hamamelis (Hamamelidaceae); Siphini (Chaitophorinae) live on numerous genera of Poaceae, etc. Some aphid genera are confined to one or two plant families, e.g., Uroleucos feed only on Asteraceae and Campanulaceae [8,38]. Of course, such a pattern could result from the fact that many aphid taxa were circumscribed based on host association. However, these affinities of closely related aphids for particular host plant groups are confirmed by phylogenetic studies, and in some cases, aphid phylogenetic relationships are better explained by their host plant similarity than by morphological criteria. The molecular phylogenetic reconstruction of the Aphidinae clusters some aphid genera that are morphologically divergent but use related plants [36]: Myzus and Brachycaudus were thought unrelated based on their morphology [39], but nevertheless were found closely related in molecular studies, and both contain many species using Prunoideae as a host plant (as sole hosts or primary hosts). Similarly, pterocommatines (Pterocomma, Plocamaphis, etc.), feeding on Populus and Salix (Salicaceae), are so morphologically different from any other aphids as to be placed in a separate subfamily (Pterocommatinae) by a recent classification [6]; however, molecular phylogenetics placed pterocommatines firmly within Aphidinae– Macrosiphini as a sister group to Cavairellia, the only other lineage of Aphidinae that has also diversified on Salicaceae [36].

The degree of host plant specialization in phytophagous insects is difficult to define and often subjective. Depending on the breadth of their host range, aphids have been classified as monophagous, oligophagous, polyphagous and even pantophagous [24]. An aphid species that uses several plant species from the same genus probably exhibits a higher level of host specialization than a species that uses only two hosts, which are drawn from distantly related plant families. The former suggests stronger phylogenetic constraints in the colonization of new host plants and is more indicative of specialization from an evolutionary point of view. Blackman and Eastop [7,29,33] compiled an impressive body of data on host-plant association for all aphids. However, the level of host specialization in aphids has still not been extensively compared between subfamilies and genera. To date, only a few generalities have been drawn, such as all aphid species feeding on conifers are monophagous or oligophagous; aphids associated with monocotyledons are less specific than aphids associated with dicotyledons [18,40]; galling aphids (e.g., Hormaphidinae) are often very specialized [41]. Furthermore, levels of specialization in aphids have rarely been compared with other groups of phytophagous insects. Compared to other hemipterans, such as psyllids [24], aphids might not be as strictly monophagous, but they exhibit greater phylogenetic structure in their host plant ranges.

Several types of selective pressures could favour a specialized diet in aphids. Here we briefly review behavioural and life-history traits of aphids that are considered as important in shaping host-plant association ([19,42] for more comprehensive discussions). Plant chemical defences (secondary compounds), which are usually involved in plant-insect coevolution [43–45], have rarely been implicated in shaping long-term patterns of aphid-plant associations. As phloem feeders, aphids probably circumvent many of those defences ([9] pp. 32–38 regarding the occurrence of chemical defences in phloem). One hypothesis to explain their specialist diet is simply that particular aphid species just perform better on particular plants that are more suitable as a resource for them; specificity would thus be the result of metabolic adaptation (see section 3) [42]. Mechanisms (behavioural, chemical) of host-plant location and recognition in aphids may also play an important role in either favouring or discouraging specialization [9]. However, the fact that some aphids host alternate between very different plant species argues against host specificity being solely a consequence of adaptation for finding and using the resources offered by a specific plant. An alternative hypothesis has been proposed by Ward [46], who suggested that host specificity was favoured because host plants serve not only as a source of nutrients but also as a mating site (termed the “rendezvous” hypothesis [9]); hence, host-range limitation is probably strongly linked to mating success. This hypothesis is clearly in agreement with the observation that heteroeocious species exhibit a wider diversity of secondary hosts and a narrower range of primary hosts, where the sexual phase of the cycle is produced and mating occurs. Of the 141 heteroeocious European Aphidinae, only nine have their secondary host plants confined to one genus [24]. Another strongly debated hypothesis is that of “fundatrix specialization” [37,47,48]. Under this hypothesis, host specialization in aphids is due largely to the phylogenetic constraints on the morph that hatches from overwintering eggs in spring, the
fundatrix, and perhaps also the sexual morphs. These morphs would be strongly adapted to the primary, often ancestral hosts, and less able to switch hosts than would other morphs. This hypothesis is also consistent with the general pattern of a narrower primary host-plant range than secondary host plant range in heteroecious species.

2.3. Phylogenetic studies of the role of host plants on speciation modes in aphids

There are no studies on the evolution of host association at the level of Aphididae as a whole; however, several case studies have focused on the historical association of certain aphid tribes and genera and their host plants in order to investigate how specialization, cospeciation and host shifts have influenced diversification. These studies are summarized in Table 1 and reviewed below.

Through a phylogenetic reconstruction of North American Cinara, Favret and Voetglin [49] proposed that speciation was accompanied by host shifts. Closely related species in the Cinara phylogenetic tree are often found associated with different Pinus species, while they maintain similar feeding sites (they are either branch or shoot feeders). This suggests that these aphids are better at switching hosts than at colonizing different ecological niches on the same host plant. Although the Pinus species serving as hosts of closely related Cinara are sympatric, no information exists on whether these host shifts and specialization occurred in sympatry or whether they were facilitated by geographical separation of populations.

Several phylogenetic and ecological studies of galling aphids in Fordini have been published [50–52]. Fordines are often host specific and repeated host shifts between related hosts have been inferred [51]. However, several occurrences of aphid sister species sharing the same host plant species suggest that host shifts followed by host specialization and speciation are not the sole mechanism promoting diversification in this group. Study of two sympatric Pemphigus species sharing the same host (P. populi and P. obesinymphae on Populus deltoide) suggests that allochronic speciation (temporal isolation) has occurred. Hence, variation in plant phenology could be an important factor in generating aphid diversity [53].

Containing more than 197 valid species, Uroleucon constitutes one of the many genera that probably arose during the diversification of aphids on herbaceous plants in the Tertiary. Most Uroleucon species are generally restricted to a single host plant species or to several closely related species within Asteraceae or Campanulaceae. In some cases, species complexes are associated with groups of closely related plants; thus, Uroleucon is a seemingly good candidate for an example of aphid-host parallel diversification. However, a molecular phylogenetic reconstruction of the genus, based on only 17 species, clearly does not support a cospeciation hypothesis [38]: Uroleucon species have shifted repeatedly between Asteraceae tribes. Furthermore, the phylogeny implies that Uroleucon ancestors were not associated with a common ancestor of Campanulaceae and Asteraceae, as would be expected under a cospeciation scenario; rather, the association with Campanulaceae is a derived condition. Estimation of the age of the radiation suggests that Uroleucon species have diversified after Asteraceae became abundant in the fossil record, and speciation linked to host specialization has more likely originated through the acquisition of existing host taxa. Again, it is difficult to conclude whether host specialization followed by speciation could have occurred in sympathy based on molecular phylogenies alone. The authors of this study mentioned that most host plants of Uroleucon overlap over large parts of their geographic range; therefore,

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Sampling</th>
<th>Host plants</th>
<th>Patterns of host plant association</th>
<th>Putative speciation mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cinara (Lachninae)</td>
<td>25 sp./150 sp. in the genus [49]</td>
<td>Mainly Pinus, but also Picea, Abies, Juniperus, Pseudotsuga</td>
<td>All species specialized on a genus of Conifers, 16 species specialized on one or two species, 9 more generalists. The same conifer species often hosts several Cinara sp.</td>
<td>Host switch</td>
</tr>
<tr>
<td>Tribe Fordini</td>
<td>16 sp./70 in the tribe [51]</td>
<td>Pistacia and various secondary hosts</td>
<td>Specialized on one or two species of Pistacia. No specialization on secondary hosts. The same Pistacia sp. often hosts several Fordini</td>
<td>Host switch</td>
</tr>
<tr>
<td>Pemphigus (Eriosomatinae)</td>
<td>Phylogeography of 2 sp. and phylogeny of 8 sp./70 sp. [53]</td>
<td>Populus</td>
<td>Most species are specialized on one or several Populus sp. The same Populus sp. can host several Pemphigus sp.</td>
<td>Allochronic speciation</td>
</tr>
<tr>
<td>Uroleucon (Aphidinae)</td>
<td>15 sp./180 sp. [38]</td>
<td>Asteraceae, Campanulaceae</td>
<td>Most species are specialized on one or several plant species within the same genus, several polyphagous species</td>
<td>Host switch</td>
</tr>
<tr>
<td>Cryptomyzus (Aphidinae)</td>
<td>9 sp./17 sp. [54]</td>
<td>Ribes and/or Labiatae</td>
<td>Specialized on one or several species of Ribes and/or one or several plant species within Labiatae</td>
<td>Loss of host alternation Host switch</td>
</tr>
<tr>
<td>Brachycerus (Aphidinae)</td>
<td>27 sp./50 sp. [56]</td>
<td>Prunus and/or various families of herbaceous plants</td>
<td>Specialized on one or several species of Prunus and/or specific to various species of herbaceous hosts, several polyphagous species</td>
<td>Loss and gains of host alternation Host switch</td>
</tr>
</tbody>
</table>
because mating does occur on the specific host plant, reproductive isolation could be favoured even in sympathy. Rapid diversification (as indicated by the lack of resolution in the deep nodes of the phylogeny) is also in favour of a sympatric mode of speciation [38].

A thorough study of speciation mode in Cryptomyzus aphids was conducted using phylogenetic studies and experimental work [54,55]. There are 17 valid species in this genus, with some species alternating between Ribes and several herbaceous species of the Lamiaceae family, and other species being monoeocious on Lamiaceae. Two possible pathways of speciation were proposed in Cryptomyzus: host shift of a monoeocious species onto a new herbaceous host followed by specialization and reproductive isolation (similar to the scenario inferred for Uroleucon [38]), and speciation via loss of the heteroeocious life cycle and shift of all primary-host generations onto some of the secondary hosts of the original life cycle [54,55].

The genus Brachycaudus presents a pattern of life cycle evolution resembling the situation in Cryptomyzus. It includes heteroeocious species (often alternating from a Prunus tree to herbaceous hosts from very diverse angiosperm families) and monoeocious species that are often found on the secondary hosts of closely related heteroeocious species [56]. Brachycaudus species also vary greatly in their host plant repertoire, ranging from strictly monophagous (most monoeocious species) to highly polyphagous on secondary hosts when heteroeocious; the latter condition might favour the capture of new hosts. This genus potentially illustrates speciation via acquisition of new, sometimes distantly related, hosts during the “summer” phase of the heteroeocious life cycle, and through changes in life cycle (Jousselin et al. in prep).

Overall, the few studies of the history of host association within aphid genera all demonstrate that speciation is often associated with a host shift. However, these macroevolutionary patterns do not firmly demonstrate the role of plant use on reproductive isolation in aphids. At the most, they provide hints that host plant specialization, together with life cycle evolution, has probably played central roles in aphid diversification. Assessing the role of host plant association as a promoter of speciation necessitates intraspecific studies within natural populations, which are reviewed below.

3. Host specialization and divergence in natural populations: How can host plants drive aphid speciation?

3.1. Ecological speciation and patterns of host association

Host plants exert considerable selective pressures on their associated aphids, constituting their only food resource and habitat (including mating and oviposition sites). Because the vast majority of aphid species are host-specific (see section 2), it is tempting to infer their origin and divergence as a direct consequence of adaptation to different host plants, a hypothesis that was proposed long ago for phytophagous insects [57]. However, distinct host specificities may just result from local adaptation and competition between species that have previously diverged in geographical isolation (allopatry) [58]. Under such processes, host specialization would merely allow the coexistence of host-adapted insects as a consequence of reproductive isolation, rather than causing it. This hypothesis does not appear to be considered in the scientific literature on aphid ecology [9].

By contrast, speciation promoted by divergent selection in contrasting environments, a process termed “ecological speciation” [59], is actively studied in a variety of organisms, including fishes, birds, fungi and plants [59,60]. Premating isolation (reduction of hybridization rate) may arise from mating occurring within the preferred habitat. In addition, the contrasting parental environments may not be suitable for hybrids, thereby causing post-zygotic isolation (hybrid unfitness). While the contribution of divergent natural selection to particular components of reproductive isolation has been quantified only in a handful of biological models [61–63], phytophagous insects, including aphids, offer good systems for studying these ecological reproductive barriers [64–66]. Ecological speciation can theoretically occur under any geographical context [67], but its study requires populations in partial reproductive isolation, in which ecological and non-ecological reproductive barriers can be distinguished (this would not be possible if no hybrid is produced). In phytophagous insects, these populations would correspond to host races [64].

Few insect species have been shown to encompass host-specialized populations undergoing appreciable hybridization (typically at a rate greater than 1%) while maintaining significant genetic differentiation [64,68], such as the well-known races of the apple maggot fly, Rhagoletis pomonella [69–71]. Within this scarcity of host race systems, there is still a non-negligible proportion of aphids, comprising attested or serious candidate models, most belonging to the most diverse subfamily, Aphidinae. The pea aphid, Acyrthosiphon pisum that essentially comprises monoeocious populations (reviewed in reference [72]) encompasses at least eight conspecific host races specialized on different legume species [73], embracing the much-studied “clover” (Trifolium sp) and “alfalfa” (Medicago sativa) races in Europe [74,75] and North America [76,77]. In Hyalopterus pruni, two populations using either plum (Prunus domestica) or almond (P. dulcis) as primary hosts are highly genetically distinct in Europe and North America (a zone of introduction), but they hybridize on a shared host, apricot (P. armeniaca), at a significant rate (approximately 4%) [78]. The polyphagous aphid Myzus persicae comprises a specialized form on tobacco (Nicotiana tabacum) as a summer host [79] that is formally designated as the subspecies M. persicae nicotianae [7]. In Greece, sexual populations in tobacco-growing areas are genetically differentiated from those in other regions [80]. Distance alone cannot explain this differentiation, as the tobacco-feeding population is genetically uniform in most of the country. Reproductive isolation is likely ongoing, as suggested by behavioral studies of mating between tobacco lineages and those feeding on other crops, which occurs on the shared primary host, peach (Prunus persica) [80].

Given the moderate genetic differentiation between host-associated populations (populations analyzed at many microsatellites do not form genetic clusters as distinct as in the host race models presented above), biotypes of M. persicae probably have not reached species status, and
may constitute host races in areas of sexuality and sympathy.

In several other aphid species, host races are suspected but not yet confirmed due to unknown hybridization rates, or even to a lack of any population genetic analysis. Here, the distinction of “biotypes”, often named as subspecies, denotes any degree of population divergence below or at the species level [64] and the presence of minute morphological variations (body color or the presence of wings in specific morphs), sometimes confirmed on the basis of performance and preference assays on different host species. Müller [11,81] reviewed complexes of biotypes belonging to monoecious species, such as *Aulacorthum solani*, *Acyrthosiphon malvae* and * Macrosiphum euphorbiae*. Each complex presents various subspecies of unclear status, with various degrees of sympathy, host specificity and ability to hybridize. The *Aphis fabae* complex encompasses at least three subspecies alternating between *spindle* (*Euonymus europaeus*) and various herbaceous plants, and a fourth monoecious subspecies that remains on the primary host (*spindle*) all year round (case reviewed in reference [82]). These subspecies show distinct summer host preference and some degree of behavioral isolation [83,84], but most of them can be hybridized in the laboratory [82]. A comparable system is found in *Pemphigus bursarius* (*Pemphigidae*), whose specialized populations on distinct Asteraceae are highly genetically differentiated despite reproducing on a shared primary host, poplar [85]. The case of *Cryptomyzus galeopсидis* is opposite, in that two proposed host races (within a larger complex) colonize different primary hosts, *redcurrent* (*Ribes rubrum*) and blackcurrant (*Ribes nigrum*), but share their summer host (case reviewed in reference [64]). The absence of mating preference in the laboratory and the identification of possible hybrid clones [86] suggest host race status, but hybridization in the field is limited by preference to different primary hosts [87].

In addition to typical aphid populations that alternate asexual and sexual generations on a regular basis (see Simon et al., this issue), we may mention a number of host-associated biotypes comprising asexual lineages, and sometimes single clones. These occur in *Schizaphis graminum* on cereals [88], *Theroaphis trifolioli* on clover and alfalfa in Australia [89] and *Aphis gossypii*, which presents specialized clones on unrelated host families (cotton, potato and Cucurbitaceae, amongst others) [14]. In the following, we do not consider asexual biotypes, as host plants cannot promote their reproductive isolation.

Hence, intraspecific differentiation onto different host plants is not unusual in aphids, at least in the Aphidinae. Moreover, many host-associated biotypes that do not differ morphologically might have gone unnoticed.

### 3.2. Host plants as catalysts of reproductive isolation in aphids

#### 3.2.1. Sources of divergent selection and post-zygotic isolation

**Feeding performance.** In many complexes of biotypes, host specialization is measured through reciprocal transplants between host plants under laboratory conditions. Under such a controlled setup, the cause of pronounced differential performance most likely entails nutrition in the broad sense: sap intake and assimilation, and reaction to associated plant defenses. Among the physiological and biochemical factors known to affect aphid nutrition [90,91], none has been unambiguously linked to food specialization in related biotypes. For instance, amino-acid requirements differ among pea aphid races [92], but performance on artificial diets designed to mimic the amino acid composition of the phloem sap from the different hosts does not reflect performance on the plants themselves [93].

Perhaps more relevant to speciation are the genetic components of differential performance, which few studies were able to distinguish from the effect of conditioning, since lineages of different biotypes are generally reared on their respective native plant before fitness experiments. Conditioning (including maternal effects) can markedly improve feeding performance on alternate hosts, without totally explaining difference observed between lineages from different plant origins ([94,95] for tests on *Sitobion avenae* and *Aphis gossypii*, respectively). In pea aphids, performance on alternate hosts is only marginally altered by conditioning [96] and strong specialization to the native host is still observed after rearing on a shared host plant prior to tests [73,97].

Genetically based food specialization may induce reproductive isolation if intermediate genotypes, typically hybrids, present mitigated performance, as the expression of a fitness tradeoff preventing the optimal use of both parental environments. Such ecological hybrid unfitness is difficult to distinguish from non-ecological causes of poor performance, mainly intrinsic genetic incompatibilities affecting hybrids independently of their environment [98]. A rare example comes from a cross between pea aphid lines from red clover and common broom (*Cytisus scoparius*) where resulting hybrids performed very poorly on the parental plants, but reportedly grew well on the universally favorable broad bean (*Vicia faba*) [99]. The quantitative genetics of performance in hybrids between races on red clover and alfalfa revealed genomic regions conferring a higher fecundity on a parental host species, but a lower performance on the other [76]. This genetic fitness trade-off prevents the evolution of a generalist genotype on both red clover and alfalfa and explains the intermediate hybrid performances on these plants. Mackenzie [100,101] also demonstrated a trade-off between the optimal use of broad bean (*Vicia faba*) and nasturtium (*Tropaeolum majus*) by different genotypes of *Aphis fabae* and F2 crosses. In other models, crosses between host-specialized biotypes gave disparate results (reviewed in references [11,81]). Some hybrid lines can complete their cycle normally on at least one parental host, suggesting the absence of strong intrinsic genetic incompatibility. Others show various deficiencies in the laboratory, such as poor egg hatching success and defects of sexual morph production [102], which are unlikely to be related to nutrition or host adaptation in general. The strength of post-zygotic isolation often corroborates the perceived degree of divergence between biotypes, considered as races or good species, but it does not constitute evidence for a contribution of host plants in this reproductive
barrier. However, it seems reasonable to propose that feeding on several host plants is subject to trade-offs and that it may be a common source of hybrid unfitness in aphids.

**Phenology and life cycle.** An insect’s life cycle may not be optimal on several plant species with contrasting phenologies (reviewed in reference [66]). In aphids, trade-offs might prevent the optimal timing of some crucial phases of the life cycle on hosts that are suitable for different periods. Egg-hatch is genetically based and synchronized with spring budburst in some species [103, 104]. In *Aphis spireicola*, two biotypes on mikan (*Citrus unshiu*) and *Spiraea thunbergii* primary hosts differ by approximately a month in the dates of fundatrix emergence [105]. Hybrid lineages emerging before bud break on mikan starve to death [106], providing a possible case of ecological selection against gene flow through maladapted duration of egg dormancy, in addition to probable food specialization [107]. Similar divergent selection on adult emergence time may act on biotypes of *Tetaneura yezoensis* producing galls on *Ulmus davidiana* or *U. laciniate* [103], but selection against hybrids has not yet been investigated. Overall, divergent selection by host phenology in aphids may not be as universal as selection by food resource. Perennial herbaceous hosts of monoeocious biotypes, like most of those constituting the pea aphid complex [108], may be favorable all year round and would not strongly select for different parasite life cycles.

Aphids may be under divergent selection by other features of host plants and their associated habitat, such as morphology, host-associated enemies and mutualists (reviewed in reference [9]), but few of these have been investigated in relation to aphid speciation. Notably, in aphids as in many sap-sucking insects, bacterial symbionts might influence nutrition [109, 110]. The obligate endosymbiont of Aphididae, *Buchnera aphidicola* supplements the imbalanced phloem sap by supplying aphids with essential amino acids [111]. Specific symbiont strains are equivalent to cytoplasmic loci affecting feeding performance (see above). However, some facultative bacterial symbionts, which are frequent in aphids, may be acquired by lateral transfer [112], and may be partly seen as elements of the aphid ecological niche affecting host adaptation. So far, there is little evidence for the influence of secondary symbionts on host use by aphids. In one pea aphid clone, the suppression of the bacterium *Regiella insecticola*, prevailing in pea aphids from clovers (*Trifolium*), critically reduced performance on *Trifolium repens* while performance on *Vicia sativa* remained unchanged [113]. This result has not been reproduced in other clones or in other symbionts that are associated with different host plants of the complex [114–116]. The association of facultative symbionts with specialized populations or clones is not general [117] and it is not clear how the presence of symbionts may act against hybrids, the same way a genetic tradeoff in host performance can.

3.2.2. **Host plants as a source of premating isolation**

Many aphid biotypes feed on different plant species that grow within their dispersal range. Mating on these different hosts translates into habitat isolation (reduction of encounter rate between populations mating within distinct habitats). The relevance of habitat isolation to speciation assumes the incompleteness of other reproductive barriers. In *Hyalopterus pruni* and *Acyrthosiphon pisum*, occasional hybridizations between races meeting on the same host plants have been observed [73, 78], suggesting that premating isolation in the field can largely be enhanced by specialization to distinct hosts.

The first premating barrier expressed in the aphid life cycle, during plant colonization, is preference to distinct host species. The stages and influencing factors of host plant selection in aphids are well described [91], but those explaining differential preference among related biotypes are rarely identified. The tobacco biotype of *Myzus persicae* responds to volatile cues emitted by tobacco while the generalist biotype does not [118], indicating distinct differences in olfactory perception between them (*M. persicae* subspecies still prefer the same primary host, peach). More often, gustatory reception appears involved in differential host preference, as colonizers frequently reject alternate hosts only after probing with their stylets before reaching the phloem, under the influence of chemical stimulants or repellents [81, 91]. In pea aphid races [119], host acceptance can be triggered by epidermal extract from the home plant [120], suggesting that colonizers can sense host-specific stimulants. In this complex, differences in host preference appear genetically based and mildly affected by conditioning [81], although this observation may not apply for other specialized populations (e.g. in *Aphis fabae* [121]). Linkage between loci controlling host choice and host adaptation may further facilitate premating isolation in the face of gene flow between pea aphid races [76].

In several species, winged colonizers often show strong preference to host plants that they may not have even experienced before dispersal [122]. In heteroeocious species, returning morphs of different biotypes will prefer different primary hosts, as shown in males of *Cryptomyzus galeopsidis* mating on redcurrant or blackcurrant [87]. Because male aphids may use plant volatiles to locate mates [122] they may show particularly pronounced host discrimination. Returning females (gynoparae) also show stronger host preference than subsequent generations using the same primary host in *Aphis fabae* on spindles [123] and in *Rhopalosiphum padi* on bird cherry, *Prunus padus* [124]. These female morphs may therefore particularly enhance spatial isolation between biotypes, but their degree of preference has not been investigated in a complex of biotypes on different primary hosts.

Because plant colonization occurs before mating, host preference may actually account for most of the reproductive isolation between biotypes [125]. However, a significant proportion of colonizers choose less favorable hosts in preference assays [118, 126] or in the field [73, 127]. In this case, selection against poorly adapted immigrants may strengthen habitat isolation. In monoeocious lineages, where host colonization may long precede the mating season [9], host plants select not only against poorly adapted migrants, but also against their parthenogenetic progeny. Such selection has been evidenced in pea aphid races on red clover and alfalfa, whose genetic differentiation increases throughout the summer [12, 128].
Allochronic isolation (populations mating at different periods) is another premating barrier that could evolve between host-associated biotypes. The biotypes of *Aphis spiraecola* on mikan and *Spiraea*, which differ in emergence time (see above) also differ in the period of production of sexuals and are therefore in strong allochronic isolation [106]. The pea aphid race using the pea as a summer host is also described as producing sexuals later than the races using perennial crops [129]. Allochrony in sexual morph production could match plant phenologies, but no causal link has been strictly demonstrated. In all likelihood, the contribution of allochrony to premating isolation in host-associated aphids is probably weaker than the contribution of host preference and selection against immigrants, which affect virtually all sympatric biotypes mating on different host species.

In biotypes sharing the same primary host, specialization to different secondary hosts cannot directly promote premating isolation. However, summer host plants may indirectly promote premating isolation between biotypes on the same primary host if they select against hybrids (see section 3.2.1). The avoidance of maladaptive hybridization, termed reinforcement [130], may explain behavioral isolation in several complexes of biotypes [80,131]. Females emit pheromones at different times of the day corresponding to periods of male activity in *Myzus persicae* [80] and *Aphis fabae* [83] on the peach and spindle, respectively, and between *Cryptomyzus galeopsidis* cryptic species on redcurrant [131,132]. However, the hypothesis that these premating barriers have evolved by reinforcement remains to be tested in these models.

### 3.3. Evidence and context of speciation through host shifts

The mechanisms by which host plants may induce reproductive isolation in aphids (Table 2) constitute various aspects of host specialization. However, widespread host specialization in aphids may actually have evolved as a consequence of divergence and does not constitute evidence for a direct role of host plants in speciation. In only few models have some specific predictions of ecological speciation been verified (Table 2, bottom). An experiment by Shaposhnikov on *Dysaphis* [133–135] (Table 2) might still constitute the most remarkable demonstration of host-induced speciation. In *Dysaphis anthrisci* (feeding on *Anthriscus*), Shaposhnikov successfully transferred one clone to *Chaerophyllum*, a normally unsuitable plant that hosts the related species *D. chaerophyllina*. In about 20 parthenogenetic generations, the *D. anthrisci* clone apparently adapted to its new host, gained some morphological resemblance and reproductive compatibility with *D. chaerophyllina*, while almost losing all ability to produce viable eggs with its parent species! This unique case of experimentally repeated, or reversed, ecological speciation (depending on the original direction of the host shift in nature) should however be taken with great caution, as the possibility of contamination of the reared *D. anthrisci* clone by some *D. chaerophyllina* line has not been excluded.

### Table 2

Observations relevant to host-induced speciation in aphids.

<table>
<thead>
<tr>
<th>Category</th>
<th>Observation</th>
<th>Biological models (non exhaustive list)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Features favourable to host-induced speciation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ecological specialization</td>
<td>Range restricted to few related plant species or genera</td>
<td>Most Aphididae [24]</td>
</tr>
<tr>
<td></td>
<td>Intraspecific specialization to distinct host species</td>
<td>Many Aphidinae (reviewed in references [11,64,81,82])</td>
</tr>
<tr>
<td>Life cycle</td>
<td>Feeding resources (selective agents) are mating sites</td>
<td>All aphids</td>
</tr>
<tr>
<td></td>
<td>All stages can feed on the same host plant/species</td>
<td>All non host-alternating biotypes</td>
</tr>
<tr>
<td>Host-induced post-zygotic isolation</td>
<td>Trade-offs in performance on the parental hosts affect hybrids</td>
<td><em>Acyrthosiphon pisum</em> [76], <em>Aphis fabae</em> [100]</td>
</tr>
<tr>
<td></td>
<td>Maladapted life cycle (timing of egg hatch) in hybrids</td>
<td><em>Aphis spiraecola</em> [105,106]</td>
</tr>
<tr>
<td>Host-induced premating isolation</td>
<td>Preference to favourable mating sites (hosts)</td>
<td><em>Acyrthosiphon pisum</em> [119,126], <em>Cryptomyzus spp.</em> [87]</td>
</tr>
<tr>
<td></td>
<td>Selection against immigrants from other host species</td>
<td><em>Acyrthosiphon pisum</em>, [12]</td>
</tr>
<tr>
<td></td>
<td>Timing of production of sexual morphs may match the host phenology</td>
<td><em>Aphis spiraecola</em> [106], <em>A. pisum?</em> [129]</td>
</tr>
<tr>
<td></td>
<td>(possible allochronic isolation)</td>
<td><em>Myzus persicae</em> [80], <em>Aphis fabae</em> [83], <em>Cryptomyzus galeopsidis</em> [131]</td>
</tr>
<tr>
<td></td>
<td>Behavioural isolation on shared primary hosts</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(possible reinforcement)</td>
<td></td>
</tr>
<tr>
<td><strong>Tests of host-induced reproductive isolation (RI)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is RI lower in the absence of selection by host plants?</td>
<td>Hybrids perform better on non-selective medium</td>
<td><em>Acyrthosiphon pisum</em> [99]</td>
</tr>
<tr>
<td></td>
<td>Substantial hybridization is observed on shared hosts or despite limited migration</td>
<td></td>
</tr>
<tr>
<td>Does the strength of RI positively correlate with that of divergent selection?</td>
<td>Differences in host adaptation do not correlate with levels of genetic differentiation between biotypes</td>
<td><em>Acyrthosiphon pisum</em> [73], <em>Hyalopterus pruni</em> [78]</td>
</tr>
<tr>
<td>Does adaptation to different hosts induce genetic differentiation?</td>
<td>Loci controlling host use induce widespread genomic differentiation between biotypes</td>
<td><em>Acyrthosiphon pisum</em> [140]</td>
</tr>
<tr>
<td>Can RI be induced by selection on different host plants?</td>
<td>Adaptation to a new host induce RI with lineages maintained on the ancestral host</td>
<td><em>Dysaphys anthrisci?</em> (see text) [133–135]</td>
</tr>
</tbody>
</table>
The geographical context of divergence may also indirectly suggest an active role of host adaptation, if prolonged allopatry is excluded. Sympatric speciation in animals is a hotly debated topic [125,136] that has long been proposed in phytophagous insects, especially after the work on *Rhagoletis pomonella* host races [70]. This mode of ecological speciation assumes that host-associated populations remain within each other’s dispersal range and that gene flow, starting at 50% (panmixia), is maintained until speciation is over. In practice, allopatric divergence cannot be excluded in aphids, as known species complexes have wide continental ranges, and because geographical distributions are not thoroughly known. For instance, certain European countries do not seem to harbor some subspecies of *Aphis fabae* [11], suggesting some possible geographic separation. Biotypes of *Hyalopterus pruni* occurring in sympatry across their widely studied range are weakly geographically differentiated [15], but their origin of several million years ago leaves ample time for earlier allopatry. On the other hand, the *Myzus persicae nicotianae* subspecies may present a rare case of divergence onto a recently introduced crop, tobacco (cultivated in Eurasia since the sixteenth century), similar to the emergence of the *Rhagolepis pomonella* race after apple was introduced in North America. Given the extreme polyphagia and wide distribution of *Myzus persicae* s.s. [7], strict allopatric divergence of the subspecies *nicotianae* appears improbable. Similarly, the very recent divergence of pea aphid races and species may have been promoted by the increasing availability of some of their legume hosts through agriculture and climate warming in the Holocene [137]. It is not clear how such a context would favor repeated allopatric divergences on plants that are now broadly sympatric, given that the radiation occurred after the last glacial episode. Furthermore, the eleven pea aphid biotypes show variable rates of hybridization and genetic differentiation that may illustrate the gradual evolution of host races into good species [73].

4. Conclusion and future prospects

Patterns of aphid–plant associations at recent or deeper evolutionary timescales are compatible with widespread plant-driven diversification in the aphid family, with the major radiation of Aphididae following the diversification of angiosperms. Future developments in molecular phylogenetics should focus on obtaining reliable evolutionary relationships of Aphididae subfamilies as well as investigating their monophyly in order to confirm major transitions in aphid-host plant associations. Current taxonomic and phylogenetic studies within subfamilies and genera reveal that divergence events typically involve shifts to new plant species that are often related to the ancestral host, and occasional changes in life cycle from host alternation to monoecy and vice versa. More studies investigating the history of host plant association within aphid genera should allow a better understanding of these patterns. Due to their pest status, aphids have generated an impressive amount of data on host plant association and thorough taxonomic studies. Global analyses such as comparative studies of levels of host plant specialization and diversification rates in aphids represent promising perspectives for elucidating the role of host plants in aphid diversification.

Divergent natural selection imposed by several components of the host-plant niche, in particular food resource and phenology, may promote pre- and post-zygotic reproductive barriers, permitting speciation in the absence of complete geographical isolation. However, the hypothesis of diversification actively driven by host plants has been seldom tested with specific predictions (Table 2), and must be weighed against scenarios where adaptation to different host plants would be a consequence rather than a cause of reproductive isolation. This overlooked hypothesis suffers from the use of host plant information to delineate related aphid species or biotypes. Studies conducted above the species level cannot discriminate between non-ecological factors and the influence of host plant on reproductive isolation. However, population studies may allow comparisons between levels of reproductive isolation among populations with different degree of ecological divergence (same host plant vs. different host plants), but with similar divergence time [62]. Ecological speciation predicts higher reproductive isolation in ecologically different populations. Under a similar approach, correlating the strength of reproductive isolation with the degree of ecological divergence between populations [138,139] may reveal whether host adaptation generally contributes to divergence in aphids. Experimental studies may also focus on reproductive isolation between biotypes under various setups, in order to compare its strength in the presence and absence of selection by host plants. Finally, the production of discrete morphs under different environmental conditions (termed as polyphenism) is an intrinsic characteristic of aphids that could be investigated in the framework of speciation. The timing of production of sexual and winged migrating morphs, for instance, shapes life cycles in connection with host phenology and host alternation. Loci controlling this trait may be a target of divergent selection and contribute to reproductive isolation between aphid populations on different host ranges.

Acknowledgments

We thank J.-L. Bonnemain for giving us the opportunity to present this review, and two anonymous referees for their valuable comments on the manuscript.

References


