

CHAPTER FIVE

LINEAGE CO-DIVERSIFICATION: ARYTAININE PSYLLIDS AND THEIR LEGUME HOSTS

Abstract

To evaluate the extent of phylogenetic congruence between a group of phytophagous insects and their host plants, psyllid (Hemiptera) and legume (Leguminosae) phylogenies are compared. Inferred cospeciation events are tested by applying an absolute time scale to both the plant and insect phylogenies. The time scales are estimated using a nonparametric rate smoothing method, calibrated on the maximum geological age of the youngest Canary Islands (La Palma and El Hierro; 1-2 Myr). A comparison of the plant and psyllid phylogenies suggests that, whilst rare cospeciation events may play a significant role in promoting diversification, systemic host switching complicates historical reconstructions of psyllid-legume interactions. Psyllids appear to be opportunistic specialists with host switching occurring when the plant lineage fluctuates in geographical abundance, population structure or through dispersal. Although host switching may be frequent, psyllid preadaptation to particular hosts is evident in many cases, and selection of a new host may be constrained by plant chemistry and architecture. Thus, successful establishment by a psyllid colonist is likely to be more common when available hosts are phylogenetically and ecologically related to the original host. A history of parallel cladogenesis between psyllid and legume lineages is rejected in favour of a fluctuating lineage model of co-diversification which presents a more realistic interpretation of the present day pattern of host associations.

5.1 Introduction

It has frequently been noted that herbivorous insects may show strong patterns of taxonomic association with their hosts plants (Mitter, Farrell & Futuyma, 1991; Thompson, 1994). However, when cladograms of host and herbivore groups are compared the associations are frequently complicated and show relatively little cospeciation (parallel cladogenesis). The program TreeMap (Page, 1994a) which analyses host-parasite interactions, is designed to

maximize cospeciation events when reconciling host and parasite trees. While this is appropriate for situations in which cospeciation is the dominant process, it may not be suitable for 'messy' data sets which have comparatively little cospeciation relative to other types of interaction between host and parasite lineages. Cospeciation is only one of the processes generating patterns of coevolutionary associations, and localized coevolution (both temporal and spatial) may take place between lineages despite historical incongruence (Thompson, 1994). Host switching onto related plants is another process that can possibly maintain taxonomic associations without parallel cladogenesis. I examine here the evidence that systemic host switching in a highly non-random fashion could be a mechanism which explains the observed patterns of strong taxonomic association in plant-herbivore interactions. A more realistic interpretation of the evolution of plant-herbivore associations may be addressed by a fluctuating lineage model described below.

5.1.1 *The psyllid-legume system*

Psyllids (Hemiptera, Sternorrhyncha) are sap-feeding insects which complete their life cycle on a single host. They are host specific and widespread on dicotyledonous angiosperms (a few species feed on monocotyledons, and over wintering adults can be found on gymnosperms). One subfamily of psyllids (Arytaininae) contains three closely related genera which are found exclusively on brooms (tribe Genisteeae, Leguminosae). The Leguminosae (pea family) comprise one of the largest and most economically important families of flowering plants. Within the legumes, the Genisteeae (brooms, gorse and relatives) are a group of, usually shrubby plants, particularly well represented in the Mediterranean.

The three genera of broom-feeding arytainine psyllid (*Arytaina*, *Arytainilla* and *Livilla*) are considered to be a monophyletic group (c. 90 known species) (Chapter 3) which appear to have diversified in parallel with the diversification of the genistoid legumes (c. 260 species). These three genera are only found on genistoid host plants and are most often restricted to a single host species. The Genisteeae and their associated psyllids are well represented in the Canary Island archipelago. A detailed survey of the ecology and biogeography of the island species allows the evolution of host-parasite associations to be studied against a background of island radiation. These radiations can then be dated using the geological (K-Ar based) dates of the islands (Ancochea *et al.*, 1994; Guillou *et al.*, 1996), as has been done in a study of diversity in Canary Island beetles (e.g. Emerson, Oromí & Hewitt, 2000).

5.2 Methods

5.2.1 *Tree construction*

Phylogenetic trees were produced using PAUP* (Swofford, 1999) from nuclear (ITS1-5.8S-ITS2) data for legumes and mitochondrial (12S rRNA, and cytochrome oxidase: COI-tRNA-COII) and morphological data for the psyllids, using the same methods as previously described (Chapters 3 & 4). For the legumes, 11 taxa were added from published GenBank sequences (accession numbers given in Table 1) to those previously obtained (Chapter 4). The equally weighted MP searches recovered three equally parsimonious psyllid trees and ten equally parsimonious legume trees. For each group (psyllid and legume) a single tree (identical to one of the equally weighted MP trees) was recovered on successive reweighting (using mean rescaled consistency indices), and these tree topologies were used in subsequent analyses. Although maximum likelihood (ML, molecular data only) for these data and the MP trees were not identical they were broadly congruent.

5.2.2 *Rate smoothing*

The reweighted parsimony topologies were used with the ML branch lengths estimated using a ML model, the parameters of which were determined by a nested hierarchical method using Modeltest (Posada & Crandall, 1998). The best model for both data sets, based on Akaike Information Criterion (AIC), was general time reversible with invariable sites and gamma distribution. The ITS branch lengths were used for the legumes and CO branch lengths were selected for the psyllids (see below). As the molecular clock was rejected for all data sets (based on the χ^2 likelihood ratio test with and without the molecular clock enforced, i.e. evolutionary rates are variable across both legume and psyllid lineages), the branch lengths were rate-smoothed using the nonparametric (NPRS) method of Sanderson (1997) as implemented in TreeEdit (Rambaut & Charleston, 2000). The CO branch lengths were used in preference to the 12S branch lengths for the psyllid tree because the increased amount of data in the longer CO region improves the accuracy of the NPRS method (Sanderson, 1997). The rate smoothed trees were age-calibrated using geological dates for the islands of El Hierro and La Palma. These islands were used because they have a single estimated maximum age. The particular node selected for each tree was based on the presence of the most divergent

taxon/clade endemic to La Palma or El Hierro. The selected node was assigned with the maximum age of the island (El Hierro, 1.1 Myr for the legumes; and La Palma, 2 Myr for the psyllids). Estimates of the error for the NPRS dates was established using 100 randomly weighted bootstrap replicates. For each bootstrap replicate the branch lengths were estimated for the given psyllid and legume trees, and the distribution of the resulting ages for each node and its standard deviation were calculated using the program r8s (Sanderson, 1997).

5.2.3 *Tanglegram*

In order to estimate the amount of cospeciation, the program TreeMap (Page, 1994a) was used to compare psyllid and legume phylogenies using reconciled trees (Page, 1994a, 1994b) (Fig. 1). A heuristic search (default settings) and the proportional-to-distinguishable randomization test (1000 randomizations) were implemented.

5.2.4 *Associated nodes*

Inspection of the trees indicated that individual nodes on the psyllid tree could be associated with nodes on the legume tree. In order to determine correspondence between psyllid and legume nodes, psyllid nodes were optimized onto the legume tree using the program MacClade (Maddison & Maddison, 1992) with an accelerated transformation (ACCTRAN). Each psyllid clade was mapped onto the legume tree and the legume node corresponding to this psyllid clade was then recorded (Fig. 2 and Table 2). However, where there is incongruence between psyllid and legume trees, it becomes more difficult to determine the corresponding nodes. I used a default optimization method which associated the psyllid node, with the basal legume node for the present day host or host clade. Where the psyllid node could be optimized to more than one legume node, the oldest legume node was selected. In some cases this led to a nonlinear sequence of legume node ages associated with a linear sequence of psyllid ages (Figs 2 & 4). The hypothetical chronology of psyllid and legume trees was determined by rate smoothing the molecular ML estimated branch lengths on the MP topology (Fig. 2). This allowed corresponding legume and psyllid nodes to be dated. The dates of the associated nodes are compared in Figure 3.

5.2.5 *Field observations*

Host specificity was determined by a wide geographical survey in which all relevant legume species in the Canary Islands and adjacent areas were examined for the presence of nymphs

and adults (nymphs are apterous, and hence are good indicators of the correct host plant). In all, c. 315 legume populations were examined. Of these, c. 295 had associated psyllid populations. Egg placement and nymphal feeding sites were noted. In most cases samples were made across the geographical range of the host plant, and for many species sampling covered different times of the year in three consecutive years, in order to obtain information on seasonal fluctuations in psyllid populations (see Chapter 2).

5.3 Results

5.3.1 *Tanglegram*

The ‘tanglegram’ (Fig. 1) shows a complex pattern of host associations. Little parallel cladogenesis is apparent. However, the randomization test in TreeMap indicates that the 15 cospeciation events (suggested by TreeMap, and indicated by black circles on the nodes in Fig. 1) were significantly greater ($P = 0.005$) than the number expected by chance, and five additional cospeciation events (marked in red in Fig. 1) were recovered when secondary host associations were discounted. Secondary host associations may obscure cospeciation events if psyllids are present on a host only in the absence of the primary psyllid parasite. These results suggest that cospeciation is an important element of psyllid-legume interactions. It should, however, be noted that TreeMap maximizes cospeciation events by tracing the insect-host association back to the most recent ancestral node, and I detail below reasons why this method results in some of the 15 cospeciation events being spurious.

5.3.2 *Cospeciation and host switching*

There may be many possible reconstructions of the historical associations between host and parasite when there is incongruence between the phylogenies. Examining the pattern of host associations in terminal sister taxa may provide the best available evidence of historical evolutionary processes. At the tips of the tree there is good evidence for some cospeciation, where sister species of psyllid occur on sister host species (e.g. A, B on Fig. 1). However, many of the cospeciation events suggested by the TreeMap model appear to be spurious (e.g. C, D on Fig. 1). These are nodes in the psyllid phylogeny which represent relatively recent speciation events between psyllids that occur on phylogenetically divergent hosts (see Fig. 2). The reconciliation method implemented in TreeMap traces the host-parasite association back

to the most recent common host ancestor, which, as the divergence of the hosts increases, will be mapped to increasingly deeper nodes in the host phylogeny (i.e. young psyllid nodes will map to much older legume nodes). It is more likely that recent host switching to relatively distant host species accounts for the incongruence between psyllid and legume phylogenies. The dynamics that precipitate host switching are interesting because the majority of host legumes are associated with a single psyllid species, if host switching is common one might expect multiple psyllids on a single host. Field data on the conditions under which host sharing by psyllids can occur is therefore important (see below).

5.3.3 Optimization and dating of nodes

Figure 3 indicates that the majority of the psyllid nodes are markedly younger than their associated legume nodes in this reconstruction (i.e. the majority of comparative points in Fig. 3 fall below the diagonal and are therefore not contemporaneous cospeciation events). The relatively younger psyllid nodes suggest that the general psyllid-host association pattern is not the result of cospeciation but that psyllid colonization is sequential to the origin of the legume hosts.

Where the phylogenies are incongruent the optimization of the psyllid nodes onto the legume tree may result in a nonlinear age sequence for the corresponding legume nodes. This occurs in the event of a host switch from a recently derived host to an evolutionarily older host, in which case the default optimization associates the psyllid node with the basal legume node of the present day host. A more realistic placement of the psyllid node on the legume lineage would be at some point in time further up the branch of the present day host, that would postdate the divergence of the host group on which the ancestral psyllid occurred (this is illustrated in Fig. 4 – B, C and D). Shifting the time of host colonization up the host branch is therefore required to linearize the age sequence of associated host nodes after host switching.

Another example of discordant age sequences between legume and psyllid nodes is evident where the divergence of the psyllid taxa appears to predate the divergence of their present day hosts. A markedly earlier divergence of psyllids relative to their legume hosts is only found in the ‘Macaronesian clade’ (Table 2, and Figs 2 & 3). The Canarian psyllids provide an isolated endemic group which has an estimated minimum colonization age of 4.8 Myr. The estimated minimum colonization age for the primary host group (*Teline*) in the Canaries is only 3.8 Myr. This age discrepancy (1 Myr) for the initial diversification of the

island legume and psyllid lineages may not be significant as the standard deviation for the NPRS dates for all nodes in both psyllid and legume trees is 0.2-0.9 Myr. However, four additional examples in the ‘Macaronesian clade’ exhibit a greater discrepancy between an earlier psyllid evolution and later host evolution. Two of these examples involve psyllid species which are near the base of the ‘Macaronesian clade’ (*A. sp.5* and *A. sp.6*, 2.2-4.6 Myr) but they occur on more recently evolved host species (0.7-2.4 Myr, Fig. 2, nodes 27 & 28; and Fig. 4C). The other two examples involve sister psyllid taxa which occur on sister host taxa (Fig. 2, nodes 35 & 36; and Fig. 4A), but the divergence between the psyllids is > 2 Myr, and the host taxa have identical sequences for the ITS region, giving 0 Myr for the time of host divergence.

As the occurrence of psyllid divergence prior to their host plants is only evident in the ‘Macaronesian clade’, there may be elements responsible for this pattern that are peculiar to the evolution of the island legume-psyllid associations. On the one hand the psyllids may have evolved and diverged prior to colonizing present day hosts, and then switched to the current hosts without speciating. The plausibility of this scenario depends on how closely related the ancestral hosts were to the present day hosts. If the current and ancestral hosts were distantly related (e.g. *Teline* and *Adenocarpus*), the absence of speciation during the switch to the current host is unlikely, based on the absence of such dual host associations in present day species. Moreover, the independent colonization by sister psyllid taxa of sister hosts would appear to be an unlikely coincidence. A second explanation involves the erosion of genetic divergence between the island host group due to hybridization and introgression, or via introduction of genetic material from more recent colonizations or human introduction. The presence of genetic exchange by these processes may be sufficient to obscure the original age of host divergence.

5.4 Discussion

5.4.1 Determinants of host specificity in psyllids

Evidence from field data, and from the psyllid and legume phylogenies, suggest that plant characteristics (most likely chemistry and surface properties) are important in restricting the host choice of psyllids. The *Adenocarpus* species are extremely abundant in the Canary Islands but have only been colonized by psyllids once. Their chemistry (presence of certain

pyrrolizidine and biperidine alkaloids) and glandular surface make them unlike other Canarian legumes. *Chamaecytisus*, on the other hand, has a plant architecture that is superficially similar to *Teline*, and has been colonized four times (see Chapter 3). Three of the colonizations of *Chamaecytisus* were apparently from *Teline*, and were within the same small clade of psyllids which may have been preadapted to make this host switch. The extraordinarily high diversity to area ratio of psyllids in the Canary Islands may partly be explained by the creation of vacant niches each time a legume colonizes a new island. The pattern of psyllid distribution on *Adenocarpus* hosts provides evidence that the presence of a psyllid on a legume prevents colonization by other psyllids (presumably by competitive exclusion). The current distribution pattern is as follows: on Tenerife *A. nigrilineata* is restricted to, and abundant on, *A. foliolosus*; whilst *A. proboscidea* is restricted to, and abundant on, *A. viscosus*. Under these conditions *A. proboscidea* is never found on *A. foliolosus* except in the extensive host hybrid zones. On La Palma, however, only one psyllid is present, *A. proboscidea*, and although it is more abundant on its primary host here, it can also be found on *A. foliolosus* (which, on La Palma, lacks its associated psyllid from Tenerife).

5.4.2 Determining factors of psyllid occurrence on hosts

Field survey data indicate that rare legumes (even from clades otherwise associated with psyllids) tend to lack psyllids. For instance, *Genista benehoavensis* is a rare species, presently with > 1000 individuals, but in the 1970s there were less than 50 individuals recorded (Santos-Guerra, 1975). *G. benehoavensis* lacks a psyllid but its sister taxon on Madeira, *G. tenera*, is a common species and has an associated psyllid which is part of a clade of *Teline*-feeding psyllids, which suggests that the phylogenetic switch from *Teline* to *Genista* is not prohibitive, but insufficient host abundance may have prevented *G. benehoavensis* from being colonized. In another example *Teline gomerae*, *T. nervosa* and *T. pallida* (< 2000 individuals each) lack psyllids, although this legume clade includes the host species *T. rosmarinifolia* and *T. splendens*, with > 3000 individuals each. No legume rarer than 2000 individuals has an associated psyllid, implying that psyllid extinction may occur if a host population drops below a certain level. However, population fragmentation may also be important, e.g. *T. rosmarinifolia*, which occurs in several small populations of c. 500-1000 individuals, and at the time of sampling the associated psyllid was a rare presence in only one of the two populations sampled. Conversely, extremely abundant hosts appear to be very favourable to psyllid association. On Tenerife, *T. canariensis* (with > 10,000 individuals) occurs in several

habitats and climatic zones and has four associated psyllids – *A. pileolata* is abundant throughout the range of *T. canariensis*, but the other species tend to be geographically and ecologically restricted. Other common and widespread hosts in the Canary Islands with more than one psyllid are *Teline stenopetala* (with four psyllids) and *Chamaecytisus proliferus* (with three psyllids), and continental examples of the same pattern include *Adenocarpus complicatus* and *Cytisus scoparius* (each with two psyllids).

5.4.3 *Fluctuating lineage model*

Analysis of the phylogenetic trees as well as field data, indicate that host switching occurs throughout the diversification of the two groups (i.e. it is systemic). It is also non-random, with host switches to related hosts more typical. The situation in *Adenocarpus* indicates that host switching is unlikely if there is a psyllid already in residence on the host. However, a vacant niche may be exploited initially by a poorly adapted psyllid which gradually becomes selected to suit the host chemistry, phenology and surface characteristics. Evidence of how vacant niches may arise is taken from field data showing the absence of psyllids on rare hosts. If a legume species goes through a bottleneck (resulting in extinction of the psyllid fauna) or a population is established by long distance dispersal without its psyllid (as would be the case in island colonization), it may then recruit a host-switching psyllid when a population expands sufficiently to support a psyllid fauna. Thus, although a lineage may originate with an autochthonous (native) psyllid inherited from its ancestor, it may end up with an allochthonous (foreign) psyllid – though typically from a related legume lineage – thus breaking the parallel cladogenesis pattern.

The evidence presented here suggests that psyllids colonized the Genisteae early in the history of the group, so there is the potential for every broom to have a psyllid, inherited without host switching, from this ancestral association. However, the Genisteae lineage is clearly a fluctuating environment for psyllids and, where host numbers in a lineage drop below a certain level, its psyllid may be lost. If the lineage becomes common again it is an available niche for a psyllid, and can be colonized by a psyllid effecting a host switch. I call this the fluctuating lineage model of co-diversification. However, colonization of a new host is more likely to succeed if the switch occurs from a related host (preadaptation in the psyllid colonist). This model is shown diagrammatically in Figure 5, and may be a more realistic approach to the complex patterns of plant-herbivore associations. Further studies are needed to

determine whether there is evidence for this model from other plant-herbivore associations (such as phytophagous lepidoptera; Menken, 1996).

5.4.4 Concluding remarks

Can historical associations be reconstructed given present distributions? If the hypothetical psyllid and host phylogenies derived from the reconstruction in Figure 5 are analyzed with TreeMap, the resulting TreeMap reconstructions under represents the number of host switches (maximum two host switches in TreeMap, but seven in Figure 5) and over represents the number of cospeciation, duplication (parasite speciation without host speciation) and sorting events (parasite extinction). However, a history of opportunistic host switching and stochastic extinction (such as that illustrated in Figure 5) may never be recoverable, and phylogenetic interpretations may remain susceptible to misinterpretation of the historical associations.

5.5 References

- Ainouche A-K, Bayer RJ. 1999.** Phylogenetic relationships in *Lupinus* (Fabaceae: Papilionoideae) based on internal transcribed spacer sequences (ITS) of nuclear ribosomal DNA. *American Journal of Botany* **86**: 590-607.
- Ancochea E, Hernan F, Cendrero A, Cantagrel JM, Fuster JM, Ibarrola E, Coello J. 1994.** Constructive and destructive episodes in the building of a young oceanic island, La Palma, Canary Island, and genesis of the Caldera de Taburiente. *Journal of volcanology and geothermal research* **60**: 243-262.
- Emerson BC, Oromí P, Hewitt GM. 2000.** Colonization and diversification of the species *Brachyderes rugatus* (Coleoptera) on the Canary Islands: evidence from mitochondrial DNA COII gene sequences. *Evolution* **54**: 911-923.
- Guillou H, Carracedo JC, Pérez Torado F, Rodríguez Badiola E. 1996.** K-Ar ages and magnetic stratigraphy of a hotspot-induced, fast growing oceanic island: El Hierro, Canary Islands. *Journal of volcanology and geothermal research* **73**: 141-155.
- Käss E, Wink M. 1997.** Phylogenetic relationships in the Papilionoideae (family Leguminosae) based on nucleotide sequences of cpDNA (*rbcL*) and ncDNA (ITS 1 and 2). *Molecular Phylogenetics and Evolution* **8**: 65-88.

- Kunkel G. 1976.** Biogeography and Ecology in the Canary Islands, eds. The Hague: Dr. W. Junk Pub., .
- Maddison WP, Maddison DR. 1992.** MacClade. Analysis of Phylogeny and Character Evolution. Version 3. Sunderland, Massachusetts: Sinauer Associates.
- Menken SBJ. 1996.** Pattern and process in the evolution of insect-plant associations: *Yponomeuta* as an example. *Entomologia Experimentalis et Applicata* **80**: 297-305.
- Mitter C, Farrell B, Futuyma DJ. 1991.** Phylogenetic studies of insect-plant interactions: insights into the genesis of diversity. *Trends in Ecology and Evolution* **6**: 290-293.
- Page RDM. 1994a.** Parallel phylogenies: reconstructing the history of host-parasite assemblages. *Cladistics* **10**: 155-173.
- Page RDM. 1994b.** Maps between trees and cladistic analysis of historical associations among genes, organisms, and areas. *Systematic Biology* **43**: 58-77.
- Posada D, Crandall KA. 1998.** Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**: 817-818.
- Rambaut A, Charleston M. 2000.** TreeEdit. Version 1.0-a4.61 (program distributed by the authors). Oxford: University of Oxford.
- Sanderson MJ. 1997.** A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution* **14**: 1218-1231.
- Santos-Guerra A. 1975.** Notas corológicas. I. *Vieraea* **5**: 89-98.
- Swofford DL. 1999.** PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sunderland, Massachusetts: Sinauer Associates.
- Thompson JN. 1994.** *The Coevolutionary Process*. Chicago: University of Chicago Press.