Testing cospeciation through large-scale cophylogenetic studies

Astrid Cruaud and Jean-Yves Rasplus

CrossMark

Insects are involved in a multitude of interactions with other organisms, which make them ideal models for large-scale cophylogenetic studies. Once phylogenies of interacting lineages have been inferred, there are a number of questions we may wish to ask, such as what was the relationship between the partners in the past? Have they co-evolved for thousands or millions of years, or has one of the partners switched among different host species? To answer such questions, researchers may conduct cophylogenetic analysis, to explore the relationships between the phylogenies of interacting lineages and determine whether the match is significant or find explanations for observed differences. When combined with dating analyses, cophylogenetic analyses may support cospeciation of the partners or phylogenetic tracking. As they may reveal dynamics of host-pathogen coevolution, cophylogenetic studies may also help tackle global health issues (e.g. document the spread of disease causing pathogens). Cophylogenetic studies of parasitoids and their insect hosts may also help identify effective biocontrol agents. With the advent of next generation sequencing technologies and keeping in mind that systematic errors may occur, cophylogenetics will benefit from better-resolved trees, allowing more accurate reconciliation. However as trees become larger, current algorithms also become more computationally challenging. Nevertheless, both theoretical and methodological developments are leading to more accurate and powerful tests of cospeciation through cophylogenetic analysis.

Address
INRA, UMR1062 CBGP, F-34988 Montferrier-sur-Lez, France

Corresponding author: Cruaud, Astrid (cruaud@supagro.inra.fr)

Insect groups involved in patterns of cospeciation

Insects are one of the most successful groups of animals on Earth. They are involved in a multitude of interactions

with other organisms such as bacteria, microsporidia, fungi, plants, nematodes, and vertebrates, which make them great models for large-scale cophylogenetic studies (Box 1) and, more generally, coevolution. Patterns of cospeciation have been observed in multiple systems that include insect lineages. Generally, cospeciation is more frequently expected when the life histories of the two interacting lineages are tightly linked such as for vertically transmitted symbionts and their insect hosts, or when interactions are species specific.

Mutualistic interactions between insects and their endosymbiotic bacteria are ubiquitous and occur in many insect groups (e.g. [1]). These interactions facilitate the use by the insects of nutritional resources from various difficult to digest sources, such as sap, wood, etc. For Hemipteran insects, endosymbionts provide nutrients to their insect hosts that feed on sap, a resource that miss essential amino acids. In these obligate interactions, congruence among phylogenies has been demonstrated (aphids and *Buchnera* [2], leafhoppers and *Sulcia/Baumannia* [3], stinkbugs (Plataspidae) and a specific gut bacterium (γ-*Proteobacteria*) that is vertically transmitted from the mother to her developing eggs [4]. Strict cospeciation also occurs between cockroaches or termites and their obligate endosymbionts (*Blattabacterium*) and/or gut microbiomes that exhibit a cellulolytic and diazotrophic activity [5,6]. Another case of mutualistic interaction in which cospeciation may be expected is mimicry (e.g. between species of *Heliconius* butterflies *H. erato and H. melpomene*). For this example, results remain unclear. The first phylogenetic analyses found contrasting histories with topological and temporal incongruence that argued against codivergence [7•]. However, using coalescent based methods and cutting-edge cophylogenetic methods, Cuthill and Charleston [8] concluded that the evolutionary history of *H. erato* and *H. melpomene* was compatible with a number of temporally congruent codivergence events.

Plant-herbivore or host-parasite coevolution can also result in patterns of cospeciation, though much less frequently. Mutualistic interactions involving insects are diverse and ecologically important [9]. In these highly specialized relationships, the two interacting species mutually benefit from their interactions. A few plant genera (*Ficus*, *Yucca*, and *Glochidion*) are exclusively pollinated by obligate seed-parasitic insects (Agaonidae wasps, Prodoxidae and *Epicephala* moths respectively). Insects pollinate the flowers and oviposit in the plant ovaries where larvae

Box 1 Glossary

When starting to compare the evolutionary history of interacting lineages, one may stumble on the following terms especially since they are frequently used interchangeably: codivergence, coevolution, cophylogeny, cospeciation. This box attempts to give what we believe are consensual definitions of these terms and others, that may appear confusing.

Coadaptation: Microevolution of two or more interacting species in response to reciprocal selection between them [49].

Codivergence: The parallel divergence of interacting lineages.

Codiversification: Correlative diversification of two or more interactive lineages or organisms. Speciation events in one lineage are correlated with speciation events in a second lineage [30•].

Coevolution: Reciprocal natural selection occurring during reciprocal evolutionary interaction between two or more organisms.

Cophylogenetics: field of research that focuses on the macro scale coevolutionary associations formed between the phylogenies of interacting lineages [50].

Cophylogeny = cophylogenetic analysis explores the relationships between the phylogenetic trees of interacting lineages. In most analyses, the goals are to determine whether the match/congruence between the two (or more) trees is significant and to find the best explanation for the differences between the trees [20].

Cospeciation: The matching of speciation events and their co-occurrence upon the time between two or more interacting lineages.

Phylogenetic tracking: A pattern in which cladogenesis occurs in parallel in two interacting lineages of organisms, but the speciation events are not synchronous (i.e. one lineage speciates first and is followed by speciation in the other).

subsequently feed on a subset of the developing seeds. In these nursery pollination mutualisms, only a few studies have investigated the level of cocladogenesis between sparsely sampled phylogenies of the mutualistic partners. Cophylogenetic analyses of yuccas and their pollinator moths showed congruence between the phylogenies, though this pattern may be better explained by biogeographic factors than by coevolution (as within a lineage, yucca species and their hosts mostly occur in allopatry) [10]. Both cospeciation and host shifts have played an important role throughout the evolutionary history of the *Glochidion* and *Epicephala* moth system [11]. Finally, the largest cophylogenetic study published so far that focussed on the fig–fig wasps mutualism [12] highlighted long-term cospeciation between the partners (Box 2).

Regarding parasites, studies have been published on ectoparasites of animals such as chewing or sucking lice that develop on the body of birds (ducks, doves, flamingos, pelecans, penguins, pigeons, seabirds or toucans) or mammals (primates, rodents). Cospeciation patterns have been demonstrated between sucking lice and heteromyid rodents [13], chewing lice and marine birds [14] and body louse and New World doves [15]. However, for most groups of lice and their vertebrate hosts,

phylogenetic congruence is not the rule. Cophylogeny between parasitoids and their insect hosts has been rarely investigated (e.g. [16]), though such studies may help to set up effective biocontrol programs (e.g. reduce unintended effects). On a more general note, most cophylogenetic studies are conducted on two-lineage systems and only a few focused on more complex systems (e.g. moth/parasitoids/plants [17]), though this may help to better understand dynamics among multiple trophic layers in an ecosystem or specialized interaction.

Analytical approaches used in cophylogenetic studies

Current methods are divided into two main groups: 'event based' and 'global fit' methods (see e.g. de Vienne *et al.* [18•] and Filipiak *et al.* [19]). Event based methods consist of mapping the 'dependent' (parasite/symbiont, etc.) phylogeny onto the 'independent' (host) phylogeny, to analyse the congruence between the pair of trees, and reconcile their shared evolutionary history. Reconciliation generally considers four evolutionary events: codivergence, duplication, host switch and loss, and supposes that parasite/symbiont, etc. may only inhabit a single host. In parsimonious event-based methods, each event is assigned a penalty score and algorithms are developed to infer a minimum cost mapping, which aims to represent the most likely shared history between the pair of phylogenetic trees [20,21•,22]. When timing information is available, software can differentiate compatible and incompatible host switches and propose alternative minimum cost scenarios for reconciliation (e.g. [23]). Obviously, event cost value, which is difficult to evaluate a priori, may impact the solution. Therefore, new methods have been developed to explore the space of cost vectors. Coevolution Assessment by a Likelihood-free Approach (COALA) tries to estimate the frequency of different evolutionary events based on an Approximate Bayesian Computation approach [24]. Starting from the phylogenies of the interacting lineages and a prior probability distribution of each event, COALA generates a number of simulated 'dependent' trees using different probabilities for each event. Trees are then compared with the known 'dependent' tree and parameter values leading to trees closest to the known tree are kept, while others are rejected. Optimal reconciliations between the two input trees are then proposed. Recent work has focused on the development of new algorithms that are not too computationally expensive to reconcile trees that become larger and larger with the advent of NGS technologies (e.g. [21•,24,25]). Algorithms have also been developed to handle the confounding effects of widespread parasites (parasites that are associated with different and sometimes phylogenetically distant hosts) [26] and to address the need for quantitative measures of phylogenetic uncertainty [24]. A major criticism of event-based methods is that comparison of