Evolutionary associations among genes, organisms and geographical areas have traditionally been studied by biologists from different disciplines, with little interaction between them. Consequently, recognition of the fundamental similarity of the problem faced by molecular systematists, parasitologists and biogeographers has been slow in coming. This is particularly true of the parallels between the relationship between gene and organismal phylogeny, and the macroevolutionary associations studied by parasitologists and biogeographers. The analogy between vicariance biogeography (organisms tracking areas) and host–parasite cospeciation (parasites tracking hosts) has been recognized for some time; for a parasite the host can be thought of as an ‘area’, hence host speciation is equivalent to a vicariance event (Fig. 1). The suggestion that these macroevolutionary patterns are analogous to the relationship between gene and species trees is a more recent development.

Types of historical association

Historical associations can be divided into three basic categories (Table 1): genes and organisms, organisms and organisms, and organisms and areas. At the molecular level, each gene has a phylogenetic history that is intimately connected with, but not necessarily identical to, the history of the organisms in which the gene resides. Processes such as gene duplication, lineage sorting and horizontal transfer can produce complex gene trees that differ from organismal trees. As such, associations between organisms, such as between hosts and their parasites (including viruses), endosymbiotes and their hosts, and insects and plants, can have a long evolutionary history, which is reflected in similarities between their evolutionary trees. At a larger scale, organisms can track geological history such that sequences of geological events (e.g. continental break-up) are directly reflected in the phylogenies of those organisms.

In each association, one entity (the ‘associate’) tracks the other (the ‘host’) with a degree of fidelity that depends on the relative frequency of four categories of events: coevolution, duplication, horizontal transfer and sorting. Consequently, each association raises the tantalizing prospect that each is a special case of a more general problem, and that a single analytical tool can be applied to all three kinds of association.

Reconstructing the history of an association

Despite the relative lack of interaction among these different disciplines, strikingly similar concepts have arisen independently from them. Parasitologists recognized the problem of multiple parasite lineages decades before Fitch’s analogous distinction between paralogous and orthologous genes. Molecular systematists and cladistic biogeographers independently developed similar methods for interpreting the history of gene trees and geographic patterns, respectively.

One implication of the parallels among the different kinds of association is that they can be studied using the same methods. Reconceived trees (Box 2) originated in molecular systematics but have been applied to both host–parasite coevolution and biogeography.

The analogies among the categories of events for the different kinds of association (Table 1) need not imply close analogy among the processes; rather, the analogy is among the patterns these processes produce. For example, although the processes of gene duplication and allele divergence are different, the resulting pattern is the same – more than one gene lineage in the same organismal lineage.

Reconstructing the history of an association

Despite the relative lack of interaction among these different disciplines, strikingly similar concepts have arisen independently from them. Parasitologists recognized the problem of multiple parasite lineages decades before Fitch’s analogous distinction between paralogous and orthologous genes. Molecular systematists and cladistic biogeographers independently developed similar methods for interpreting the history of gene trees and geographic patterns, respectively.

One implication of the parallels among the different kinds of association is that they can be studied using the same methods. Reconceived trees (Box 2) originated in molecular systematics but have been applied to both host–parasite coevolution and biogeography.

The analogies among the categories of events for the different kinds of association (Table 1) need not imply close analogy among the processes; rather, the analogy is among the patterns these processes produce. For example, although the processes of gene duplication and allele divergence are different, the resulting pattern is the same – more than one gene lineage in the same organismal lineage.
host and its associate’s phylogenetic histories have diverged. Hence, reconciled trees have been employed in two different ways: to document the history of an associate where both the host and associate relationships are presumed to be known, and to infer host relationships based on the associate phylogeny. The inference of species trees from gene trees is the paradigm instance of the latter, but there is a long history of parasitologists attempting to infer host phylogeny from parasite phylogeny and cladistic biogeographers aim to infer geological history from organismal phylogeny.

Box 1. Terminology of historical associations

Given the parallels among different kinds of historical association it is desirable to have a generic set of terms that are applicable to genes, organisms and areas. The terminology used here is illustrated below.

**Associate**: a lineage that tracks another lineage, or set of historically related entities (such as geographical areas).  
**Codivergence**: joint divergences of both host and associate. Examples include host-parasite coalescence, and vicariance.  
**Duplication**: independent divergence of the associate, with both descendants remaining associated with the host (e.g. gene duplication).  
**Horizontal transfer**: transfer of an associate lineage from one host (‘source’) to another host (‘destination’) that is not itself the immediate descendant of the source host. Examples include horizontal gene transfer and host switching.  
**Host**: the lineage or entities being tracked, such as organisms harbouring a lineage of parasites.  
**Sorting event**: an event in which an associate is distributed over only part of the host’s distribution (e.g., a patchily distributed parasite).  
**Missing the boat**: if an associate is distributed over only part of the host’s distribution, and those areas have not undergone any sorting events (i.e. it is not absent from the host), then the associate is said to be ‘missing the boat’.  
**Codivergence, Duplication, Sorting event**

<table>
<thead>
<tr>
<th>Host–associate</th>
<th>Codivergence</th>
<th>Duplication</th>
<th>Horizontal transfer</th>
<th>Sorting event</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organism–gene</td>
<td>Interspecific coalescence</td>
<td>Gene duplication, deep coalescence</td>
<td>Gene transfer</td>
<td>Gene loss, lineage sorting</td>
</tr>
<tr>
<td>Host–parasite</td>
<td>Coalescence</td>
<td>Within host speciation</td>
<td>Host switch</td>
<td>Parasite extinction, ‘missing the boat’</td>
</tr>
<tr>
<td>Organism–areas</td>
<td>Vicariance</td>
<td>Sympathy</td>
<td>Dispersal</td>
<td>Extinction</td>
</tr>
</tbody>
</table>

Table 1. Equivalent processes among different historical associations

Multiple lineages

Duplications result in multiple lineages of associates on the same host lineages. The implications of this for inferring host phylogenies have been recognized by molecular systematists dealing with gene families or multiple mitochondrial lineages. If some gene lineages go extinct or are incompletely sampled, gene trees might not faithfully reflect organismal history (Fig. 2). In contrast, parasitologists and biogeographers have (with a few exceptions) attributed discordance between host and associate trees to horizontal transfer, rather than to a combination of multiple lineages and subsequent loss of associates. Host-switching and dispersal do occur but, as they might not be the sole cause of discordance, their prevalence can be overestimated.

Box 2. Reconciled trees

The concept of reconciled trees dates from Goodman et al.’s attempts to reconcile disagreements between then-accepted mammalian evolutionary relationships and those obtained from haemoglobin genes. Largely neglected until recently, reconciled trees are now receiving renewed attention from biologists and mathematicians. Suppose we have a phylogeny for four species, and four sequences sampled from those species, and that the two trees, which we believe to be correct, disagree, as illustrated in (a).

![Reconciled tree](image)

The question is, how can the trees both be true and yet be discordant? One explanation is to embed the gene tree in the species tree (b), which requires us to postulate a number of gene duplications and subsequent losses (in this instance one duplication and three losses). This embedding can also be represented using a reconciled tree (c), which simply takes the embedded gene tree and ‘unfolds’ it so that it lies flat on the page. The reconciled tree would depict the complete history of the gene if there had been no gene losses (i.e. the three sorting events). As a consequence of the gene duplication in the ancestor of species 2, 3 and 4, we would expect those species to each have two copies of the gene. Because they do not, we must postulate three gene losses. Alternatively, the gene copies could be present but undetected. Hence, the reconciled tree makes predictions about the existence of undiscovered genes. It also suggests that genes a and c are paralogous to gene d, which is not apparent from the gene phylogeny alone.
REVIEWS

The presence of multiple associate lineages on the same hosts can lead to spurious inferences about host relationships. In this example, there are two associate lineages (a and b) on the same hosts (a). If only one associate from each host is sampled (circled in a), then it is possible to infer incorrect host relationships (b), even though the associate relationships are correct.

Applications of reconciled trees

With the increasing availability of nuclear gene sequences, reconciled trees may find ready application to the study of the evolution of gene diversity and the inference of organismal phylogeny from multiple, complex gene trees, as well as tools for database analysis. Recent work suggests that lineage sorting within a single gene may pose less of a problem for phylogenetic inference than previously thought. However, analyses at large taxonomic scales using nuclear genes are likely to encounter problems caused by gene duplication and resulting paralogy. As an example, Guigó et al. used a variant of reconciled trees to infer eukaryote phylogeny from 53 gene trees and discovered that only a third of the genes were perfectly consistent with the best fitting eukaryote tree. Although this study has flaws, it suggests that great care must be taken in using nuclear genes as phylogenetic markers – individual genes or gene families can be quite misleading about species relationships. Viewed in this light, inferring organismal phylogeny from single genes becomes as fraught as inferences based on single morphological characters, making it essential to analyze multiple genes. In the same way, parasitologists and biogeographers have stressed the need to use multiple associate trees to infer the relationships among hosts and areas, respectively.

Reconciled trees can be predictive tools. The missing lineages corresponding to sorting events (Box 2) represent associate lineages that are either extant but undiscovered, or extinct. Many sorting events in reconciled trees for genes are likely to represent undiscovered genes rather than genuine losses, given the uneven sample of sequences represented in the sequence databases. Extinct associates can also leave evidence of their previous existence. Linder and Crisp’s reconciliation of a phylogeny for Southern beech trees (Nothofagus), with a geological area cladogram, required postulating the existence of a Nothofagus clade in areas where it is currently no longer found alive but where fossils of that clade are known to occur.

‘Jungles’

Reconciled trees have nice properties but also some limitations, the most severe being that they do not accommodate horizontal transfer. Other methods, such as Brooks’ parsimony analysis (BPA), do incorporate this process, but they do not always produce biologically reasonable reconstructions. Horizontal transfer poses problems that have only recently been appreciated. Charleston has developed a solution to this problem that employs a mathematical structure called a ‘jungle’, which contains all the possible ways in which an associate tree can be mapped into a host tree, given the four processes of codivergence, duplication, sorting and horizontal transfer, and the extant associations known. Given ‘costs’ for each of these processes, it is possible to find the subgraph(s) of the jungle that corresponds to the least costly (e.g. most parsimonious) reconstruction(s) of the history of the association. This also represents an improvement in the computation time required as previous methods had to rely on heuristic procedures that were not guaranteed to find optimal solutions, whereas jungles are solved using a dynamic programming approach.

Prospects

Methods for phylogenetic analysis of historical associations are still being refined, with considerable scope for future development. The analogy between the different categories of association has proved a useful heuristic tool, but detailed analogies between the processes may prove strained. More sophisticated analyses will require careful consideration of the actual processes operating in each association, especially if maximum likelihood methods are to be developed. Alternatively, there is a case for pushing the analogy to the limit to maximize the extent to which the apparently disparate disciplines of molecular systematics, parasitology and biogeography can employ the same analytical tools.
Acknowledgements
We thank Rob Cruickshank, Richard Griffiths, Vince Smith, Faron Simon and two anonymous reviewers for their comments. This work was supported by NERC grant GR/1/895 to the first author.

References


Coming soon in TREE:

- **A crisis in the making: responses of Amazonian forests to land use and climate change, W. Laurance**
  - Chemical signals and parasite-mediated sexual selection, D. Penn and W.K. Potts
  - Sensory ecology, receiver biases and sexual selection, J.A. Endler and A.L. Basolo
  - Trees within trees: phylogeny and historical associations, R.D.M. Page and M.A. Charleston
  - Competition mediated by parasites, P. Hudson and J. Greenman
- **The adaptive significance of maternal effects, T.A. Mousseau and C.W. Fox**
  - Assessing ecosystem health, D.J. Rapport, R. Costanza and A.J. McMichael
  - Research on forest responses to high CO2, M. Jasinskiški, S.C. Thomas and F.A. Bazzaz
  - Sex determination in Hymenoptera, D. Haig
  - Action on amphibians, D.B. Wake
  - Fragments past, present and future, N. Brokaw
  - The stoop of large falcons, A. Hedenström