

RH : TEMPORAL CONGRUENCE REVISITED

TEMPORAL CONGRUENCE REVISITED:
COMPARISON OF MITOCHONDRIAL DNA SEQUENCE DIVERGENCE IN
COSPECIATING POCKET GOPHERS AND THEIR CHEWING LICE

RODERIC D. M. PAGE¹

*Department of Zoology, University of Oxford
South Parks Road, Oxford OX1 3PS, England*

Address for correspondence:
Division of Environmental and Evolutionary Biology,
Institute of Biomedical and Life Sciences,
University of Glasgow,
Glasgow G12 8QQ, Scotland, U.K.
E-mail: dpage@udcf.gla.ac.uk (publish in article)

¹ Present address: Division of Environmental and Evolutionary Biology, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8QQ, Scotland. E-mail: dpage@udcf.gla.ac.uk

Abstract.— Molecular phylogenies can be used to test hypotheses of cospeciation between hosts and parasites by comparing both cladistic relationships and branch lengths. Molecular data can also help discriminate between competing reconstructions of the history of the host-parasite association. Methods for comparing sequence divergence in hosts and parasites are described and applied to data for pocket gophers and their chewing lice. The hypothesis of cospeciation between these two clades is strongly supported. The lengths of homologous branches in the gopher and louse phylogenies are positively correlated, but little support is found for the hypothesis that lice are evolving an order of magnitude faster than are their hosts. [Cospeciation; component analysis; DNA sequences; molecular clock; phylogeny; rates of evolution; tree comparison; pocket gophers; chewing lice; *Geomys*; *Orthogeomys*; *Pappogeomys*; *Thomomys*; *Geomydoecus*; *Thomomydoecus*]

A central question in the study of cospeciation is the extent to which host and parasite cladogenesis are correlated (Brooks and McLennan, 1991). The principal test for this correlation is congruence between host and parasite phylogenies, incongruence being taken as evidence of host switching. However, equating congruence with cospeciation and incongruence with host switching is an oversimplification as incongruence may also result from the presence of multiple lineages of parasite on the same host, coupled with differential survival of those parasites (Page, 1993b; Page et al., in press). Previous work using Hafner and Nadler's (1988) allozyme data for eight pocket gophers and their parasitic chewing lice suggested that information on timing of speciation events in the two clades could be used to distinguish between these two causes of incongruence (Page, 1990b). The availability of mitochondrial DNA sequence data for 15 gophers and their lice (Hafner et al., 1994) permits reinvestigation of this question. These data also allow further exploration of the use of host-parasite systems in the study of comparative rates of molecular evolution. A general framework underlying such comparisons has been developed by Hafner and

Nadler (1990) and Hafner and Page (1995). My purpose in this paper is to apply this framework to the DNA sequence data.

MATERIALS AND METHODS

Hafner et al. (1994) obtained 379 base pairs from the same region of the cytochrome *c* oxidase subunit I (COI) gene from the mitochondria of 15 pocket gophers and their 17 parasitic lice. For details of these data and their phylogenetic implications the reader is referred to Hafner et al.'s paper. Different methods and data treatments produced slightly different trees, for convenience I have chosen to use the trees shown in Figure 1 (corresponding to host tree in fig. 2b and the parasite tree in fig. 2a of Hafner et al.). The trees are drawn as cladograms (depicting relative recency of common ancestry) and as phylograms (with branch lengths proportional to amount of evolutionary change).

In this study only substitutions at the the third codon position were considered. Branch lengths for the gopher and louse phylogenies were computed using the maximum likelihood algorithm implemented in PHYLIP DNAML 3.5 (Felsenstein, 1993). For each data set the transition:transversion (TS:TV) ratio used was obtained by employing the trees shown in Figure 1 as user trees and varying the TS:TV ratio supplied to DNAML; the TS:TV ratio that returned the highest likelihood for each data set was chosen as the best estimate of the TS:TV for that data set. All tree comparisons and randomisations described here were performed using the computer programs COMPONENT 2.0 (Page, 1993a) and TREEMAP (Page, 1995). The latter can be obtained from the Internet site <http://taxonomy.zoology.gla.ac.uk/rod/treemap.html>.

COMPARING HOST AND PARASITE CLADOGRAMS

The prerequisite for any comparison of host-parasite evolution is a reconstruction of the history of the host-parasite association. There are two principal methods for obtaining such a reconstruction: Brooks' parsimony analysis (BPA; Brooks and McLennan, 1991), and component analysis (Page, 1990a, 1994). BPA uses additive binary coding to represent parasite phylogenies, then optimizes the resulting codes on the host phylogeny. Any homoplasy is interpreted as being due to host switching or parasite extinction. Elsewhere (Page, 1990a, 1994) I have shown that BPA can produce spurious results due to the nonindependence introduced by additive binary coding, hence BPA will not be used here. Component analysis was originally developed to reconstruct biogeographic histories of taxa (Nelson and Platnick, 1981), but its similarity to Goodman et al.'s (1979) procedure for comparing gene trees and species trees suggests that component analysis may be sufficiently general to be applied to historical associations in general, including host-parasite associations (Page, 1994).

Given a fully resolved parasite tree for n taxa there are $n - 1$ internal nodes representing the speciation events in the history of the parasite clade (these are numbered 18 - 33 in Figure 1). These speciation events can be divided into three categories (Fig. 2): (1) parasite and host have speciated together (cospeciation); (2) parasite has speciated independently of the host and has remained on the host (duplication); and (3) parasite has speciated independently of the host and one of the two descendant species has colonized a new host (host switch). In addition to speciation, we need to consider the fate of parasites as they track their hosts. Parasites may go extinct, may fail to speciate with their hosts (so that the same parasite occurs on more than one hosts), may be patchily distributed so that only one of the two species arising at a speciation event has the parasite ("missing the boat"— Paterson, 1994), or may have simply avoided collection. I use the term "sorting event" to cover all four cases (Page, 1995).

Previous analysis (Page, 1990b) of the gopher-lice association using component analysis employed reconciled trees (Page, 1994), which do not (readily) allow for host switching. This defect has been addressed (Page, 1995), resulting in a method which seeks to maximise the number of cospeciation events shared between the host and parasite phylogenies. This value has well defined minimum and maximum values (0 and n , where n is the number of internal nodes in the parasite tree), and the maximum is nontrivial in the sense that simply minimising the number of host switches need not maximise the number of cospeciation events — sometimes we must postulate host switching in order to maximise cospeciation (Page, 1994, 1995). This criterion also explicitly maximises the extent to which the host tree “explains” the topology of the parasite tree (Humphries et al., 1986), and in this sense is the most economical hypothesis — postulating fewer cospeciation events requires more ad hoc explanations of the parasite tree's topology.

Comparing Cladograms

A reconciled tree (Page, 1994) is the simplest reconstruction of the history of a host-parasite association, subject to the constraint that host switching has not occurred. The reconciled tree for the gopher and louse cladograms shown in Figure 1a postulates 10 cospeciation events, six “duplications” (independent speciation of lice in situ on their host), and 27 sorting events (instances where lice are predicted to occur but do not, for example due to extinction or sampling error). Hence, of the 16 speciation events in the parasite tree, 10 can be explained as the result of cospeciation without postulating any host switching. To see whether 10 is the maximum possible number of cospeciation events given these two trees, I performed an exact search using the algorithm described in Page (1995) and implemented in TREEMAP. This algorithm tries all possible combinations of host switches, and retains those that maximize the number of cospeciation events. TREEMAP found 86 additional reconstructions with 10

cospeciation events (with between one and six host switches), but none with more.

Figure 3 shows one possible reconstruction involving a single host switch involving the louse *Geomydoecus actuosi*.

Have Gophers and Lice Cospeciated?

A simple test of the hypothesis of cospeciation is to ask whether the parasite phylogeny is independent of that of its host. If so, we would expect the number of putative cospeciation nodes shared by the host and parasite phylogenies to be no greater than that expected between the host tree and random parasite cladograms (Page, 1995). Figure 4 shows the distribution of maximum number of cospeciation events shared by the gopher tree (Fig. 1a) and 1,000 random louse trees generated by TREEMAP using a simple Markovian model (Harding, 1971) which assumes that all lineages are equally likely to speciate, and that the probability of speciation remains constant over time. None of the 1000 random trees generated share as many cospeciation events (10) with the gopher cladogram as that shared by the observed louse cladogram, hence we can reject the hypothesis that louse phylogeny is independent of gopher phylogeny ($P = 0.001$).

Choosing a Reconstruction

The gopher and lice trees are more similar than we would expect due to chance alone, but are clearly not completely in agreement (Fig. 1). Of the 16 internal nodes in the louse tree, 10 can be attributed to cospeciation with the gopher hosts, leaving six nodes to be explained by other means. The multiplicity of reconstructions found for the gopher and louse trees in Figure 1 implies that there are many possible interpretations of the incongruence between these trees. How then can we choose among these reconstructions? If a reconstruction requires one or more sorting events to be

postulated, then we might search for the missing parasites implied by those sorting events. In the case of pocket gopher lice, this is a reasonable undertaking because the number of taxa sequenced (17) represents a small fraction of the 122 described louse taxa (Hellenthal and Price, 1991).

As an example, the reconstruction in Figure 3 postulates two sorting events in the lineage ancestral to the clade of lice hosted by *Orthogeomys* gophers. Figure 5a shows this part of the reconstruction drawn as a reconciled tree (Page, 1994). This tree postulates the existence (either now or in the past) of lice related to those on *Orthogeomys*, but hosted by *Zygozomys*, *Pappogeomys*, *Cratogeomys*, and *Geomys* gophers. A morphology-based phylogeny for gopher lice (Page et al., 1995) shows lice hosted by three of those gopher genera in the predicted positions (Fig. 5b), suggesting that that this explanation is plausible. However, the disagreement between the mitochondrial DNA and morphological cladograms concerning the placement of *Geomydoecus texanus* (the DNA places this taxon with *G. ewingi* rather than as sister to the lice on *Orthogeomys*) means that further data will be needed to test the validity of this explanation for the incongruence between the gopher and louse trees.

Alternatively, we may consider choosing between the reconstructions based on the number or likelihood of the host switches and/or sorting events. The reconstructions found by TREEMAP consider solely the cladistic relationships of the hosts and parasites and not branch lengths. Some reconstructions may postulate host switches between geographically distant hosts, reducing that reconstruction's plausibility (unless the geographic distribution of the hosts has been different in the past). Another possibility is to explore the relative timing of speciation events implied by a given reconstruction (Page, 1990b, 1995). Hypotheses of cospeciation predict that putative cospeciation events in the two clades should be approximately contemporaneous. This prediction can be tested if we have some measure of the time of divergence of host and parasite.

Comparing Phylograms

Although the lice *Geomydoecus thomomyus* and *G. perotensis* are sister taxa, their hosts (*Thomomys talpoides* and *Cratogeomys merriami*) are only distantly related (Fig 1). Consider two possible explanations of this discrepancy: (1) the two lice are relicts of a larger clade whose ancestor was present on the ancestor of all pocket gophers but now is restricted to just those two hosts; or (2) *G. thomomyus* switched from its ancestral host on to *Thomomys talpoides* (equally, we could postulate that *G. thomomyus* remained on its ancestral host and *G. perotensis* colonized *C. merriami*).

If the first explanation were true, we would predict that DNA sequences from *Geomydoecus thomomyus* and *G. perotensis* are widely divergent and, if the rate of sequence evolution is similar in gophers and lice, that sequences obtained from their hosts have diverged by a similar amount. Figure 1b suggests that this is indeed the case as the common ancestor of the two lice is a relatively deep node, and hence cospeciation is a reasonable explanation of the data. Similarly, the comparable divergence between the lice *G. nadleri* and *G. trichopi* and their hosts, *Pappogeomys bulleri* and *Zygogeomys trichopus* is consistent with cospeciation. This suggests that much of the incongruence between gopher and louse phylogenies may be due to the persistence of multiple clades of lice on the same hosts. It also suggests that postulating host switches involving these lice is unnecessary, which eliminates from consideration many of the reconstructions found above. If we assume that nodes 25 and 30 in the louse tree (see Fig. 1) are cospeciation events (based on comparable amounts of divergence in parasite and host), as are nodes 18–20 and 31 (based on an exact match between host and parasite trees), then of the original 86 reconstructions only 12 remain, all involving host switching by one or more lice belonging to the clade rooted at node 26. Hence considering branch length information substantially reduces the number of reconstructions we need to consider.

COMPARING SEQUENCE DIVERGENCE

The qualitative comparisons of sequence divergence in gophers and lice made above implicitly assume that host and parasite are evolving at approximately the same rate. As Hafner et al. (1994) point out, cospeciating host-parasite assemblages offer an excellent opportunity to test the validity of this assumption, without relying on a spotty (e.g., gopher) or non-existent (e.g., lice) fossil record. Furthermore, if both host and parasite taxa possess molecular clocks we could address questions of the relative timing of host and parasite speciation (Hafner and Nadler, 1990; Hafner and Page, 1995).

Figures 6a and b show the relationship between the lengths of equivalent branches in hypothetical host and parasite phylogenies. The lack of a single, taxon-wide rate of evolution complicates our interpretation of such a plot — has parasite lineage *ii* in Figure 6a acquired less evolutionary change than its host, 2, because the parasite diverged later in time, or because it is evolving at a slower rate, or both? Answering this question requires a molecular clock.

In the case of a molecular clock all terminal taxa are equidistant from the root of the tree (Fig. 6c). This constraint means that all branch lengths are no longer statistically independent. We can restore independence by plotting only the internal branch lengths (Fig. 6d), which represents the interval in time (measured in units of genetic change) between the two speciation events that start and terminate a lineage. However, this plot loses information about relative timing of speciation events in the two clades. In the example shown, each parasite speciation precedes the corresponding host speciation by 0.5 units of time, a fact not revealed by Figure 6d. If we plot the coalescence times (i.e., the time a given pair of lineages coalesce into a single lineage) of pairs of cospeciation events (Fig. 6e), this earlier speciation by the parasites is reflected in the positive intercept of the line fitted to the points (see below).

Copaths

To compare amounts of evolution in host and parasites we need to compare homologous events in the two clades, i.e., cospeciation events. As the trees shown in Figure 6 are perfectly congruent, every internal node in those trees is a cospeciation event so that identifying equivalent branches (or coalescences) is trivial. This may rarely be the case for real data. If we want to compare coalescence times we need consider only pairs of cospeciation events, which are already given to us by the reconstruction (Page, 1994). However, identifying homologous branches requires more work.

Consider a tree in which each branch is labelled with the amount of evolutionary change that occurred along that branch. The amount of evolution that occurred between any two taxa is therefore the sum of the lengths of each branch along the path connecting the two taxa. Hence to compare host and parasite divergence we need to identify homologous paths in the two phylogenies, which I shall term *copaths*. Each member of a copath starts with either a terminal taxon or a cospeciation node, and terminates at the most recent ancestor of that node that is itself a cospeciation node (providing no host switch occurs between those two nodes). If no such node exists then there is no corresponding copath. Figure 7 shows the copaths for the gopher and louse phylogenies shown in Figure 1, based on the reconstruction shown in Figure 3. Examples of copaths include (*hispidus* – 22, *chapini* – 28) and (*trichopus* – 21, *trichopi* – 25). Some copaths may not be immediately obvious. For example, the path *perotensis* – 30 in the louse tree corresponds to the path *merriami* – 19 – 20 – 21 – 26 – 27 – 29 in the host tree, as under the reconstruction shown in Fig. 3 parasite node 30 maps onto host node 29 so that the dichotomy between *Geomydoecus thomomyus* and *G. perotensis* corresponds to the basal split within the pocket gophers.

Ideally, copaths should be statistically independent of each other and different copaths should not overlap. Unfortunately this is not always the case. Paths may

overlap in both the host and parasite tree if hosts harbour more than one parasite lineage, as occurs in *Geomys* gophers, resulting in the copaths in the parasite tree comprising nodes 22 – 24 – 26, and 23 – 24 – 26 sharing the path 24 – 26 (Fig. 7). Another source of dependence is the molecular clock, in which case evolutionary change can be represented using an ultrametric, and hence the copaths between any two terminal taxa and the same internal node will have the same length, by definition. In that case we would be better off plotting coalescence times (see above).

Is Sequence Divergence in Hosts and Parasites Correlated?

Having identified copaths (or corresponding coalescences) we can compute the correlation between copath lengths (or coalescence times) in the two clades. Is this value significant? Consider firstly the case where we have molecular clocks in both host and parasite. Taking the host phylogeny as the independent or causal variable, we wish to know whether the coalescence times for putative cospeciation events in the parasite phylogeny are independent of the coalescence times for the corresponding events in the host phylogeny. Answering this questions requires a null distribution of coalescence times. This can be obtained using various models of cladogenesis (see Nee et al., 1994b). For example, Harding's (1971) Markovian model (employed above to test the topological agreement between host and parasite cladograms) corresponds to a pure birth process with a constant birth rate, time between successive speciation events (= c

$$P(t) = \lambda e^{-\lambda t}$$

Hey (1992) described how
used to generate random coa

An alternative to a model-based null hypothesis is to generate new coalescence times by randomizing the observed times. This may be more appropriate if we have not sampled all extant members of the parasite clade, in which case accurate estimation of the parameters of birth-death processes becomes difficult (Nee et al., 1994a). Given the observed intervals between successive speciations (“horizons” in Brown’s [1993] terminology), there are $n!$ ways these can be ordered to generate new coalescence times, which can then be assigned to the parasite phylogeny using the random labelling algorithm described by Page (1991; note that the method used for generating random coalescence times in that paper assumes a uniform distribution of speciation events over time not, as stated on p. 193, a constant rate of speciation). In the absence of molecular clocks a model-based null hypothesis requires some model of character evolution to generate expected branch lengths. If a suitable model is unavailable we could again use randomisation methods, such as randomly reassigning the observed branch lengths in the parasite tree.

If sequence divergence in hosts and parasites is correlated, and both taxa have molecular clocks, then fitting a line to a plot of parasite divergence against host divergence allows us to describe two aspects of host-parasite divergence (Hafner and Nadler, 1990; Hafner and Page, 1995). The slope of the line is the relative rate of host and parasite evolution, the intercept measures the divergence of parasites when their hosts speciate. An intercept of zero implies synchronous cospeciation: host and parasite speciate together, whereas a negative intercept implies delayed cospeciation: the parasites tend to speciate after their hosts. Hafner and Nadler also noted that the intercept may be positive, in which case parasites diverge prior to their hosts such that a single host lineage may harbour more than one parasite lineage. This is analogous to the positive intercept obtained by Lynch and Jarrell (1993) when they plotted sequence divergence against taxon divergence time for a range of animal mitochondrial genes, which suggests (p. 1204) that the “mitochondria that ultimately colonize sister taxa are often substantially divergent within their ancestral species.” Parasites with restricted vagility, such as chewing lice, may show a similar pattern.

Estimating Branch Lengths

Provided that we are comparing orthologous genes, DNA sequence data allows us to compare homologous characters in both host and parasite thereby avoiding the problem of comparing non homologous morphological characters, or allozyme characters of dubious homology. With Hafner et al.'s (1994) data we can compare sequence divergence in the same region of the same gene in gophers and lice using the same units (numbers of nucleotide substitutions). In contrast, with gross morphological characters there is little, if anything, to compare in mammals and insects, making it virtually impossible to have a common yardstick for comparing amounts of evolutionary change. However, sequence data are not without their problems. Reconstructed branch lengths can underestimate actual branch lengths. Although this problem has been appreciated for some two decades (e.g., Moore et al., 1976), its implications for comparative biology have been largely ignored (but see Sanderson, 1990). Any comparison of branch lengths will need to address this issue.

If we consider the gopher and louse data, the majority of substitutions in both gopher and louse COI sequences are synonymous (silent) substitutions at the third codon position. Plotting synonymous transitions against synonymous transversions shows that transitions in both data sets are approaching saturation, more so in the louse sequences (Fig. 8). Indeed, for the lice, almost any pairwise comparison will suffer from the effects of multiple hits. Unless corrected for, this saturation will lead to underestimates of branch length, which is of course the motivation behind the plethora of distance measures developed by molecular biologists. I should also note that if the substitution process at different sites along the sequence (e.g., first, second, and third codons) differs, then the utility of a single overall measure of sequence divergence is somewhat dubious (Irwin et al., 1991).

Phylogenetic Sampling

Phylogenetic sampling also affects the accuracy of branch length estimates. The denser the sampling of lineages, the greater the chances of detecting evolutionary change (Moore et al., 1976; Fitch and Bruschi, 1987; Fitch and Beintema, 1990; Sanderson, 1990). A simple check on the adequacy of phylogenetic sampling is to plot the path length between each terminal and the root of the tree against the number of nodes along that path (Fitch and Beintema, 1990). A positive correlation, such that clades with fewer taxa have apparently accumulated fewer substitutions, is evidence that we have underestimated the amount of change undergone by sparsely sampled clades.

The COI data for both gophers and lice both show the same positive correlation ($r = 0.56$, $P < 0.05$) between path length and numbers of substitutions, suggesting that additional taxonomic sampling would be desirable. In the case of the lice we have an independent morphological phylogeny for all 122 recognised terminal taxa (Fig. 9) (Page et al., 1995) which, with the exception of *Geomydoecus perotensis* and *G. texanus*, is consistent with the molecular tree. This morphological cladogram allows us to place the sequenced taxa in a broader phylogenetic perspective, and to see that most louse taxa sampled represent single exemplars from larger clades. Gophers are also taxonomically diverse, with more than 400 recognised species and subspecies (Hall, 1981). Hence for both taxa there is considerable scope for further taxonomic sampling. Ideally, individual host clades and their parasites would be exhaustively sampled.

APPLICATION TO POCKET GOPHERS AND CHEWING LICE

Figure 10 plots the lengths of the copaths shown in Figure 7, measured as expected substitutions per third codon position. Although copath lengths are

significantly correlated ($r = 0.44$; $P = 0.01$ based on 1,000 randomizations of louse phylogeny branch lengths), the plot shows considerable scatter. It would be tempting to interpret outliers in this plot as evidence of additional host switching (in addition to that originally postulated by the reconstruction in Fig. 3), which indeed some may be. However, some care is needed when interpreting this diagram. Consider the copath (*expansus* – 26, *castanops* – 19 – 20 – 21 – 26 – 27), where the louse path has a length of 0.22 substitutions/site, whereas the corresponding gopher path has a length of 0.98 substitutions/site. The gopher path is more extensively sampled than the louse path (four clades branch off from the gopher path compared to none on the louse path), and the magnitude of the discrepancy between the two clades may partly reflect this sampling (see above). Likewise, the gross difference between the paths terminating in *thomomys* and *talpoides* (0.91 vs. 0.33) is partly an artefact of assigning the gopher lineage 28 – 29 a branch length of zero as, in the absence of an outgroup, we do not know how to partition the divergence between nodes 28 and 27 along the path 28 – 29 – 27. Were we to assign the bulk of the divergence to the ancestor of *Thomomys* (i.e., along the path 28 – 29), gopher and louse divergence would be more comparable.

Relative Rates of Evolution

Hafner et al. (1994) concluded that the lice hosted by *Orthogeomys* gophers were evolving an order of magnitude faster than their hosts. Although Figure 11 supports a higher rate for these lice, it does not indicate a 10-fold difference (reduced major axis regression of *Orthogeomys* hosted louse divergence on gopher divergence, $y = 2.61x + 0.04$). The discrepancy between this study and Hafner et al.'s is due to the different transition:transversion (TS:TV) ratio used for the lice in the two studies. Hafner et al. used a value of TS:TV = 17:1, based on comparing the most similar louse sequences. I have chosen to use the value of TS:TV that maximises the likelihood of the gopher and louse phylogenies giving rise to the observed sequence data, which for

both data sets was 4:1. Due to the saturation of transitions apparent in the COI sequences (Fig. 8), branch lengths estimated by maximum likelihood increase with increasing TS:TV, hence the magnitude of the difference between host and parasite divergence is very sensitive to choice of TS:TV (as noted by Hafner et al., 1994:1089).

Given the saturation of transitions which makes the analysis sensitive to the value of TS:TV, and the conflicting estimates of this value, I repeated the analysis using only third position transversions. The number of pairwise transversion differences between gophers and between lice were optimised onto the trees shown in Figure 1 using PHYLIP FITCH (Felsenstein, 1993). Lengths of the copaths shown in Figure 7 computed from transversions were more correlated ($r = 0.62$, $P = 0.004$) than the maximum likelihood lengths computed using both transitions and transversions. A reduced major axis regression on all copath lengths based on transversions yielded a slope of 1.28, suggesting that lice do not, on the whole, evolve substantially more rapidly than their hosts. For *Orthogeomys*-hosted lice, the reduced major axis regression has a slope of 2.65, essentially the same value computed using the maximum likelihood copath lengths. The concordance between these two estimates of sequence divergence (one with and one without transitions) suggests that the lower TS:TV ratio for the lice used here is more appropriate.

Likelihood Test

The phylogenies for *Orthogeomys* and their lice are topologically identical (Fig. 11) allowing another test of the hypothesis that the lice have diverged the same amount as their hosts, based on the likelihood difference test implemented in PHYLIP DNAML (Felsenstein, 1993). The test is that the likelihood (L) of obtaining the observed louse sequences given the louse tree topology and branch lengths shown in Figure 1b ($\ln L = -638.89$) is not significantly different from the likelihood of obtaining those sequences from the same tree topology with the branch lengths observed for the

gophers ($\ln L = -699.08$). The difference in likelihoods (-60.19) hence we reject the hypothesis of equality of sequence divergence

It is worth emphasising that the conclusion that *Orthogeomys* is evolving faster than their hosts (albeit not an order of magnitude) is dependent on the correctness of both the gopher and louse phylogenies, and on our assumptions about host-parasite associations (as well as the model of sequence evolution used, here DNAML). The clade *cavator* + *underwoodi* is only weakly supported in the louse tree. If the grouping of *underwoodi* with *cherriei* and *costaricensis* (Hafner et al., 1994) is correct, then node 20 in the louse tree is not a cospeciation event. Furthermore, in the tree for all gopher lice (Page et al, 1995), the clade containing *Geomydoecus chapini* (and its relatives on *Orthogeomys hispidus* and the *chiapensis* complexes (Price and Hellenthal, 1988) hosted by *Orthogeomys*) is supported. Sudman and Hafner's (1992) DNA sequence data place *O. grandis* as sister to the *Orthogeomys* sampled by Hafner et al. (1994), hence, if both node 20 and node 28 is not a cospeciation event either. If these deeper nodes are not cospeciation events but prior speciation events by the lice, the conclusion that *Orthogeomys* hosted by *Orthogeomys* are evolving more rapidly than their hosts is supported. Note also that other copaths, such as (*T. talpoides* – *G. barabarensis*) and (*Z. trichopus* – *G. trichopi*, *P. bulleri* – *G. nadleri*) show similar rates of sequence divergence in the two clades, and one louse copath (*T. ewingi*) has acquired considerably fewer substitutions than its host.

DISCUSSION

There are clear possibilities for using host-parasite associations to study molecular evolution of unrelated taxa (Hafner et al., 1994; Hafner et al., 1995); however such comparisons require that the history of the associations be reconstructed. Different reconstructions may have quite different

comparisons. Taxonomic sampling is also an important consideration, both for accurate estimation of amounts of molecular evolution (Fitch and Bruschi, 1987), and accurate identification of cospeciation events. In the gopher-lice association, the apparent presence of multiple louse lineages combined with the limited number (17) of louse taxa that have been sequenced (Fig. 9) results in large numbers of cladistically equally plausible reconstructions with quite different implications for the history of the association.

Once the taxonomic sampling problem is reduced, a major challenge will be to tease apart intra- and inter-clade variation in the rate of nucleotide substitution. The analysis presented here suggests that a single generalization concerning relative rates of sequence evolution in lice and gophers would be premature, but both the maximum likelihood and the transversion only branch lengths suggest that the difference of an order of a magnitude reported by Hafner et al. is open to question.

The DNA sequence data supports Hafner and Nadler's (1988) original conclusion that gophers and lice have cospeciated — gopher and louse phylogenies are more similar than expected due to chance. Furthermore, these phylogenies serve to illustrate the point that the expectation that host-parasite cospeciation will necessarily produce congruent phylogenies is naïve. If parasites speciate more rapidly than their hosts, and suffer higher rates of extinction (or fail to colonise both descendants of a host speciation event), then parasite phylogenies need not mirror closely that of their hosts, *even with little or no host switching* (Page et al., in press). This discordance will be exacerbated by incomplete taxonomic sampling. Data on seabird lice (Paterson et al., 1993; Paterson, 1994) suggests the pattern shown by the gophers and their lice is not unique. That discordant host and parasite phylogenies may be the rule, rather than the exception, even if parasites are exclusively vertically transmitted, is less surprising if we consider another class of “infectious elements transmitted exclusively from parent to offspring” (Williams, 1992:15) — genes — that can also have phylogenies somewhat divergent from that of their “hosts”.

The problem of incongruent gene and species phylogenies has implications for this study, beyond being a useful source of analogies. Although I have treated the trees shown in Figure 1 as organismal phylogenies, they are of course mitochondrial phylogenies. Patton and Smith's (1994) discovery of marked incongruence between allozyme and mitochondrial DNA phylogenies for *Thomomys bottae* and *T. townsendi* gophers is cause for concern. At the same time, it presents an opportunity for further insight into the dynamics of the gopher-lice association. It would be of great interest to know how the phylogenies of the lice found on these gophers relate to the allozyme and mitochondrial trees for these gophers (see also Nadler et al., 1990). I should note that the broad concordance between allozyme, DNA, and morphological studies of gophers and lice (Hafner and Nadler, 1988; 1990; Page et al., 1995) suggest that the trees shown in Figure 1 are indeed reasonable estimates of organismal phylogeny.

The utility of molecular data in studies of host-parasite associations extends beyond comparative molecular evolution. It may assist our efforts to unravel the various causes of congruence and incongruence. For example, although substantial topological concordance of host and parasite trees seems unlikely to be due to chance alone, it is possible that recent host switching may produce spurious congruence, especially if parasites preferentially colonize closely related hosts. Information on the amounts of sequence divergence between hosts and parasites (and /or their relative coalescence times) could help identify such causes of pseudospeciation (Hafner and Nadler, 1988). Furthermore, the hypothesis that much of the incongruence between host and parasite phylogenies may be due to differential survival of multiple lineages (Page, 1993b) predicts that parasite clades speciate more rapidly and have higher rates of extinction than their hosts. If extant members of host and parasite clades are completely sampled, these cladogenetic parameters could be estimated from their molecular phylogenies (Nee et al., 1994a, 1994b).

ACKNOWLEDGMENTS

I am indebted to Mark Hafner for sharing hard-won sequence data. For helpful comments on the manuscript I thank David Cannatella, Dale Clayton, Mark Hafner, Paul Harvey, Eddie Holmes, Jim Patton, Andy Purvis, Francis Villablanca, and Xuhua Xia. An anonymous referee provided some thought provoking comments.

REFERENCES

- BROOKS, D. R., AND D. A. MCLENNAN. 1991. *Phylogeny, ecology, and behavior*. Chicago Univ. Press, Chicago.
- BROWN, J. K. M. 1993. Probabilities of evolutionary trees. *Syst. Biol.* 43:78-91.
- FELSENSTEIN, J. 1993. PHYLIP version 3.5c. Department of Genetics, Univ. of Washington, Seattle.
- FITCH, W. M., AND J. J. BEINTEMA. 1990. Correcting parsimonious trees for unseen nucleotide substitutions: The effects of dense branching as exemplified by ribonuclease. *Mol. Biol. Evol.* 7:438-443.
- FITCH, W. M., AND M. BRUSCHI. 1987. The evolution of prokaryotic ferredoxins — With a general method correcting for unobserved substitutions in less branched lineages. *Mol. Biol. Evol.* 4:381-394.
- GOODMAN, M., J. CZELUSNIAK, G. W. MOORE, A. E. ROMERO-HERRERA, AND G. MATSUDA. 1979. Fitting the gene lineage into its species lineage: A parsimony strategy illustrated by cladograms constructed from globin sequences. *Syst. Zool.* 28:132-168.
- HAFNER, M. S., AND S. A. NADLER. 1988. Phylogenetic trees support the coevolution of parasites and their hosts. *Nature* 332:258-259.
- HAFNER, M. S., AND S. A. NADLER. 1990. Cospeciation in host-parasite assemblages: Comparative analysis of rates of evolution and timing of cospeciation. *Syst. Zool.* 39:192-204.

- HAFNER, M. S. AND R. D. M. PAGE. 1995. Molecular phylogenies and host-parasite cospeciation: Gophers and lice as a model system. *Philos. Trans. R. Soc. Lond. B* 349: 77-83.
- HAFNER, M. S., P. D. SUDMAN, F. X. VILLABLANCA, T. A. SPRADLING, J. W. DEMASTES, AND S. A. NADLER. 1994. Disparate rates of molecular evolution in cospeciating hosts and parasites. *Science* 265:1087-1090.
- HALL, E. R. 1981. *The mammals of North America*, Vol. 1. John Wiley, New York.
- HARDING, E. F. 1971. The probabilities of rooted tree-shapes generated by random bifurcation. *Adv. Appl. Probab.* 3:44-77.
- HELLENTHAL, R. A., AND R. D. PRICE. 1991. Biosystematics of the chewing lice of pocket gophers. *Annu. Rev. Entomol.* 36:185-203.
- HEY, J. 1992. Using phylogenetic trees to study speciation and extinction. *Evolution* 46:627-640.
- HUMPHRIES, C. J., J. M. COX, AND E. S. NIELSEN. 1986. *Nothofagus* and its parasites: A cladistic approach to coevolution. Pages 55-76 in *Coevolution and systematics* (A. R. Stone, and D. L. Hawksworth, eds.). Clarendon Press, Oxford, England.
- IRWIN, D. M., T. D. KOCHER, AND A. C. WILSON. 1991. Evolution of the cytochrome *b* gene of mammals. *J. Mol. Evol.* 32:128-144.

- LYNCH, M., AND P. E. JARRELL. 1993. A method for calibrating molecular clocks and its application to animal mitochondrial DNA. *Genetics* 135:1197-1208.
- MOORE, G. W., M. GOODMAN, C. CALLAHAN, R. HOLMQUIST, AND H. MOISE. 1976. Stochastic versus augmented maximum parsimony method for estimating superimposed mutations in the divergent evolution of protein sequences. Methods tested on cytochrome *c* amino acid sequences. *J. Mol. Biol.* 105:15-37.
- NADLER, S. A., M. S. HAFNER, J. C. HAFNER, AND D. J. HAFNER. 1990. Genetic differentiation among chewing louse populations (Mallophaga: Trichodectidae) in a pocket gopher contact zone (Rodentia: Geomyidae). *Evolution* 44:942-951.
- NEE, S., E. C. HOLMES, R. M. MAY, AND P. H. HARVEY. 1994a. Extinction rates can be estimated from molecular phylogenies. *Philos. Trans. R. Soc. Lond. B* 344:77-82.
- NEE, S., R. M. MAY, AND P. H. HARVEY. 1994b . The reconstructed evolutionary process. *Philos. Trans. R. Soc. Lond. B* 344:305-311.
- NELSON, G., AND N. I. PLATNICK. 1981. *Systematics and biogeography: Cladistics and vicariance*. Columbia Univ. Press, New York.
- PAGE, R. D. M. 1990a . Component analysis: A valiant failure? *Cladistics* 6:119-136.
- PAGE, R. D. M. 1990b. Temporal congruence and cladistic analysis of biogeography and cospeciation. *Syst. Zool.* 39:205-226.

- PAGE, R. D. M. 1991. Clocks, clades, and cospeciation: Comparing rates of evolution and timing of cospeciation events in host-parasite assemblages. *Syst. Zool.* 40:188-198.
- PAGE, R. D. M. 1993a. COMPONENT: Tree comparison software for Microsoft Windows, version 2.0. Natural History Museum, London.
- PAGE, R. D. M. 1993b. Genes, organisms, and areas: The problem of multiple lineages. *Syst. Biol.* 42:77-84.
- PAGE, R. D. M. 1994. Maps between trees and cladistic analysis of historical associations among genes, organisms, and areas. *Syst. Biol.* 43:58-77.
- PAGE R. D. M. 1995. Parallel phylogenies: Reconstructing the history of host-parasite assemblages. *Cladistics* 10: 155-173.
- PAGE, R. D. M., R. D. P. PRICE, AND R. A. HELLENTHAL. 1995. Phylogeny of *Geomydoecus* and *Thomomydoecus* pocket gopher lice (Phthiraptera: Trichodectidae) inferred from cladistic analysis of adult and first-instar morphology. *Syst. Entomol.* 20: 129-143.
- PAGE, R. D. M., A. M. PATERSON AND D. H. CLAYTON. in press. Lice and cospeciation: A response to Barker. *Int. J. Parasitol.* 25.
- PATERSON, A. M. 1994. Coevolution of seabirds and feather lice: A phylogenetic analysis of cospeciation using behavioural, molecular and morphological characters. Ph.D. Dissertation, Univ. of Otago, Dunedin, New Zealand.

- PATERSON, A. M., R. D. GRAY, AND G. P. WALLIS. 1993. Parasites, petrels and penguins: Does louse presence reflect seabird phylogeny? *Int. J. Parasitol.* 23:515-526.
- PATTON, J. L., AND M. F. SMITH. 1994. Paraphyly, polyphyly, and the nature of species boundaries in pocket gophers (genus *Thomomys*). *Syst. Biol.* 43:11-26.
- PRICE, R. D., AND R. A. HELLENTHAL. 1988. *Geomydoecus* (Mallophaga: Trichodectidae) from the Central American pocket gophers of the subgenus *Orthogeomys* (Rodentia: Geomyidae). *J. Med. Entomol.* 25:331-335.
- SANDERSON, M. J. 1990. Estimating rates of speciation and evolution: A bias due to homoplasy. *Cladistics* 6:387-392.
- SUDMAN, P. D., AND M. S. HAFNER. 1992. Phylogenetic relationships among Middle American pocket gophers (genus *Orthogeomys*) based on mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* 1:17-25.
- WILLIAMS, G. C. 1992. *Natural selection: Domains, levels, and challenges*. Oxford Univ. Press, New York.

Figure captions:

Figure 1. Phylogenies for pocket gophers and their chewing lice parasites, drawn as (a) cladograms and (b) phylograms with branch lengths proportional to expected numbers of substitutions at the third codon position in the COI gene (estimated using DNAML 3.5 (Felsenstein, 1993) with transition:transversion ratio of 4.0 for both clades). Pocket gopher genera are *Orthogeomys*, *Pappogeomys*, *Cratogeomys*, *Geomys*, *Thomomys*, and *Zygoeomys*; louse genera are *Geomydoecus* and *Thomomydoecus*. Internal nodes in the trees are arbitrarily numbered. Note that the phylograms are actually unrooted trees drawn as rooted trees, consequently the branches between *Thomomys* and its sister clade and between *Thomomydoecus* and *Geomydoecus*, have been arbitrarily assigned a length of zero. Thin lines connect gophers with their species specific parasite.

Figure 2. A host and parasite phylogeny showing a cospeciation event, a duplication, a host switch and a sorting event.

Figure 3. A possible reconstruction of the history of the gopher-lice association that postulates 10 cospeciation events, five duplications (in situ speciation of the lice on the same host), 20 sorting events (instances where louse lineages have been lost or remain undetected), and a single host switch (by *Geomydoecus actuosi*).

Figure 4. Distribution of the maximum number of cospeciation events shared by the gopher cladogram shown in Figure 1a and 1,000 random trees for the lice. The maximum number of cospeciation events shared by the observed gopher and louse phylogenies (Fig. 1a) is 10.

Figure 5. (a) Part of the reconciled tree corresponding to the reconstruction in Figure 3 showing the apparent absence of lice on *Zygoeomys*, *Pappogeomys*,

Cratogeomys, and *Geomys* gophers, which counts as two sorting events [O].(b) The corresponding part of the morphological phylogeny for all 122 gopher lice (Page, 1995), with the lice shown in (a) highlighted and host associations indicated by small squares. Note the occurrence of lice on *Geomys* (open squares), *Pappogeomys* and *Cratogeomys* (gray squares) as predicted by the reconciled tree (lice hosted by *Orthogeomys* are indicated by solid squares).

Figure 6. Kinds of comparison of divergence on host and parasite phylogenies. If neither host nor parasite has a molecular clock (a), we can plot the lengths of the equivalent branches in the host and parasite phylogenies (b). Given a molecular clock (c) these branch lengths are not all independent. We could plot just the internal branch lengths (d), which correspond to the intervals between successive speciation events, but information on the relative timing of speciation in the two clades is lost. This information is retained if we plot the coalescence times of each pair of cospeciation nodes (e).

Figure 7. Gopher and louse phylograms showing the copaths for the reconstruction shown in Figure 3, which postulates a single host switch by *Geomydoecus actuosi*. Each copath is a path in the host and parasite phylogenies linking pairs of corresponding nodes. By tracing out a copath in the host phylogeny one traces the host lineage(s) occupied by each parasite lineage.

Figure 8. Plots of numbers of synonymous transitions and transversions for all pairs of gopher and louse COI sequences.

Figure 9. Cladogram for all 122 described pocket gopher lice, based on morphology of adults and first instars (from Page et al. 1995: fig. 5). The subtree connecting the 17 louse for which COI sequences obtained by Hafner et al. (1994) is shown in black.

Figure 10. Plot of estimated sequence divergence at the third codon position in COI along copaths in the gopher and louse phylogenies shown in Figure 7 with a line indicating equal amounts of divergence in gophers and lice. Data points discussed in the text are labeled by the host and parasite nodes terminating the copath (i.e., that node most distant from the root of the tree). The symbols ○ and ■ indicate louse lineages parasitizing gophers belonging to *Orthogeomys* and *Geomys*, respectively; other lice are represented by solid dots. Nonindependent copaths in the louse phylogeny are indicated by square brackets.

Figure 11. Relationship between sequence divergence in *Orthogeomys* gophers and their lice (phylogenies shown in inset). Units are expected substitutions per third codon position. The regression line found using the reduced major axis method has a slope of 2.61, indicating that the lice have acquired substitutions more rapidly than their hosts. This conclusion is supported by a likelihood difference test.

Temporal Congruence Revisited

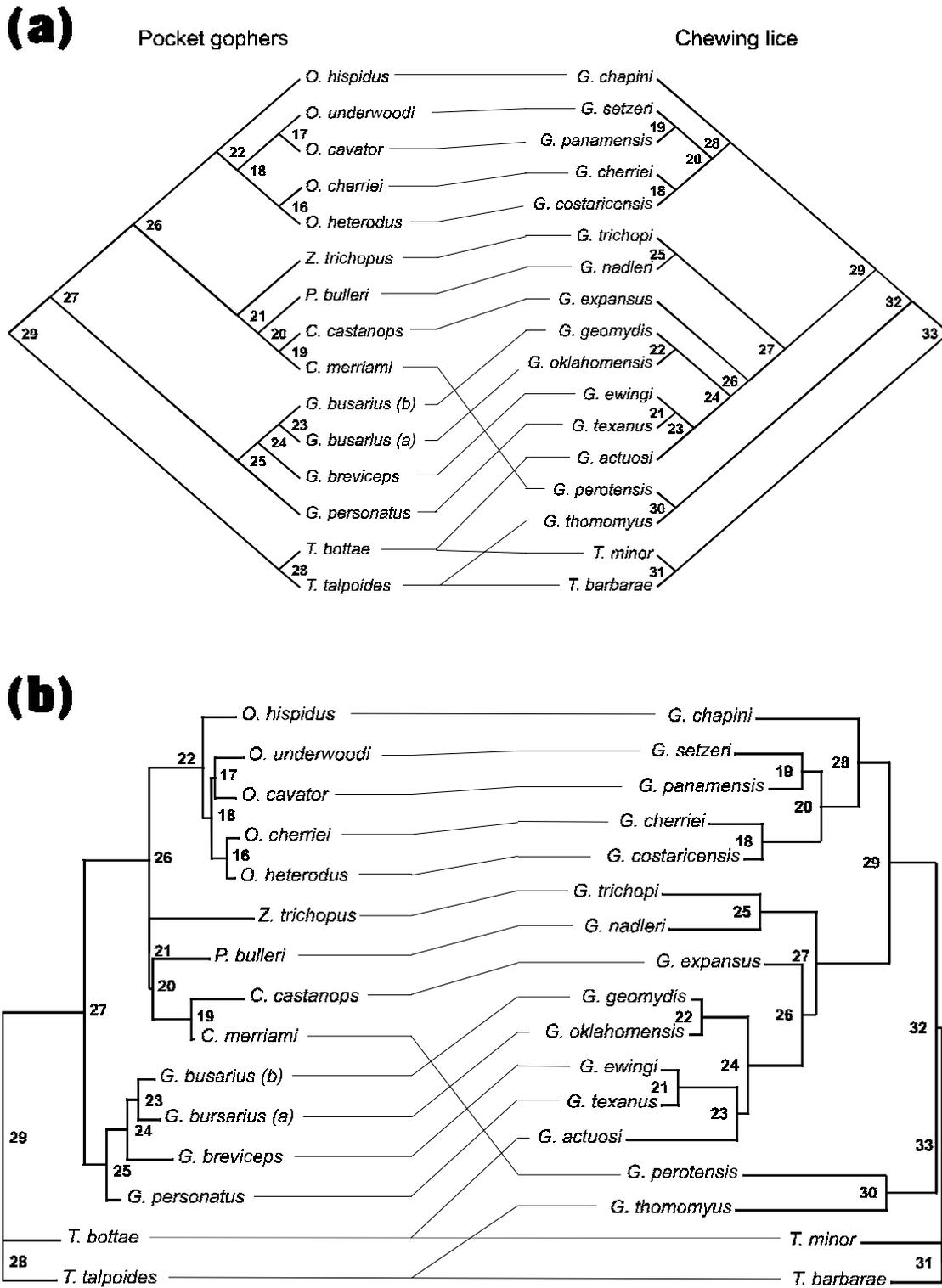


Fig. 1

Temporal Congruence Revisited

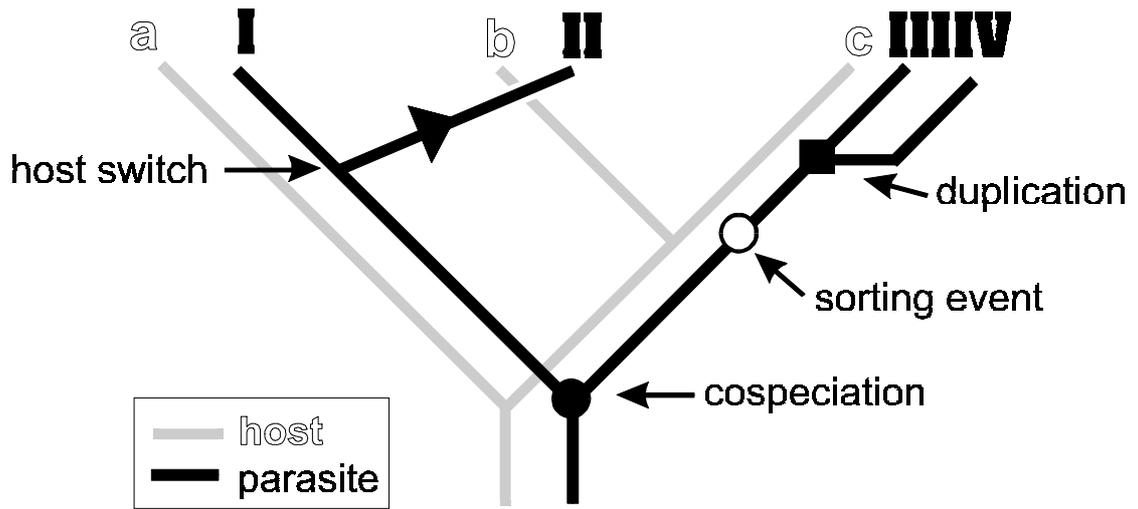


Fig. 2

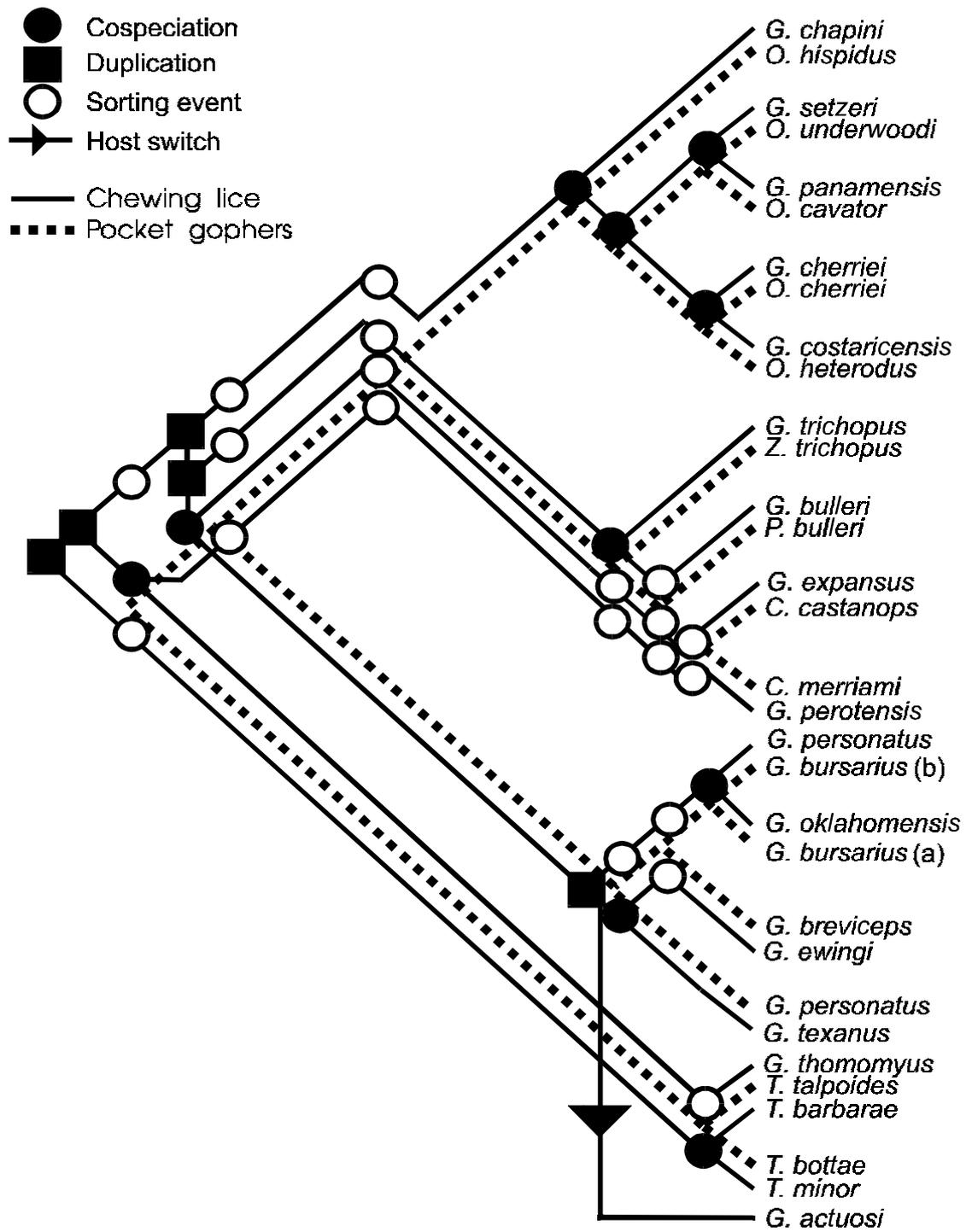


Fig. 3

Temporal Congruence Revisited

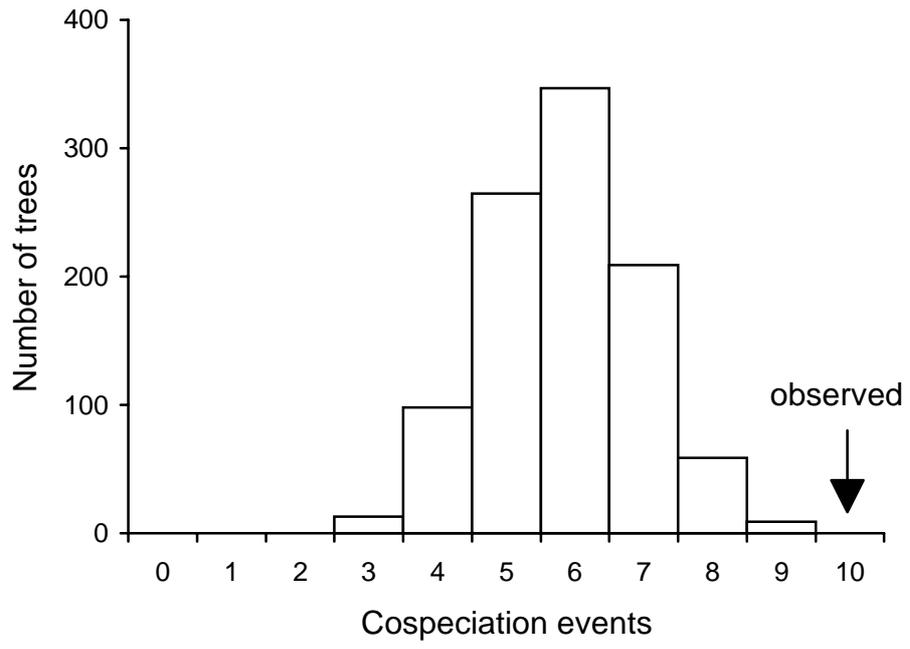
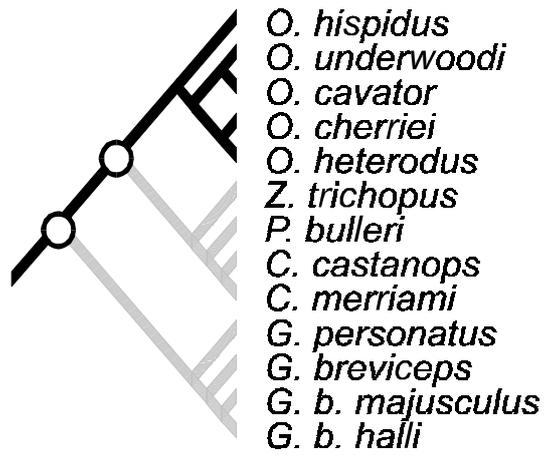


Fig. 4

(a) Pocket gophers



Chewing lice

chapini
setzeri
panamensis
cherriei
costaricensis

(b)

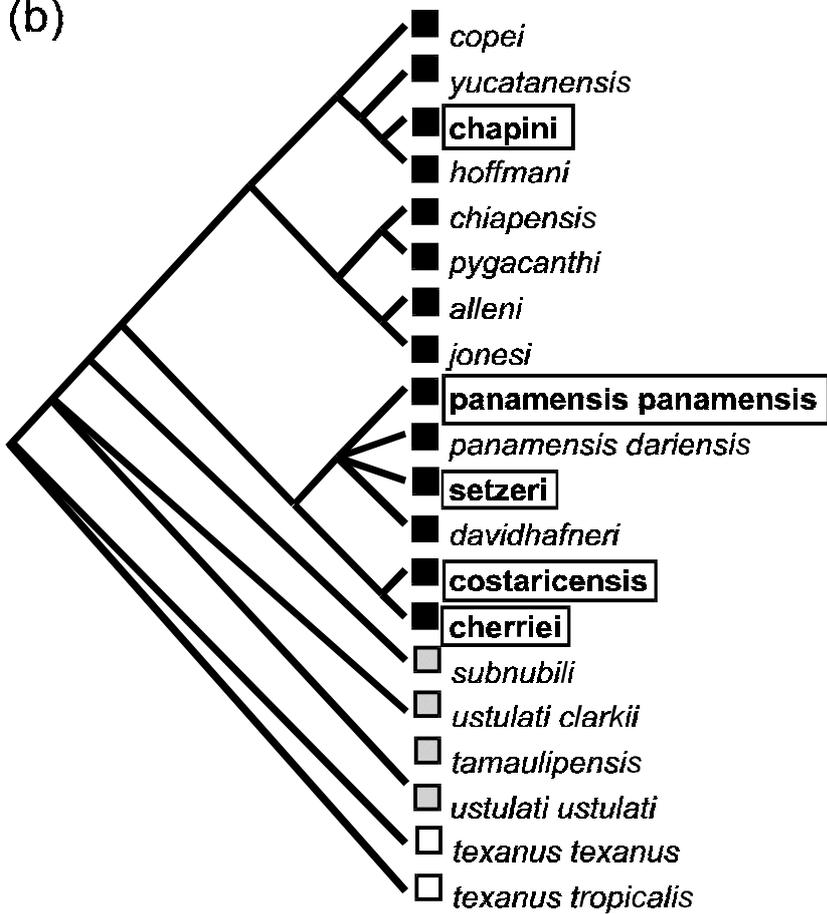
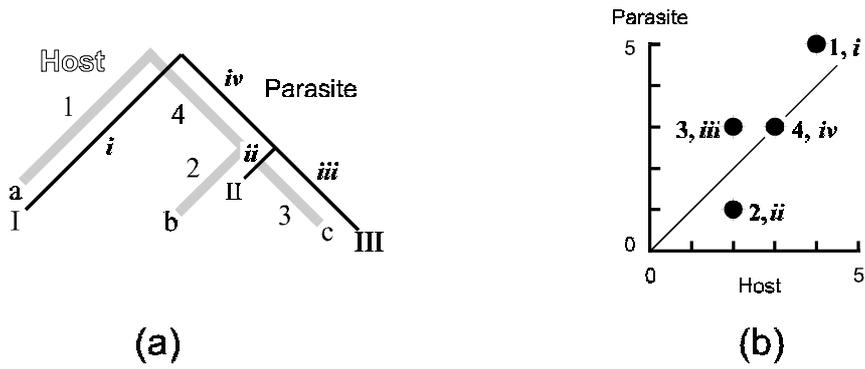


Fig. 5

Temporal Congruence Revisited

Without molecular clock



With molecular clock

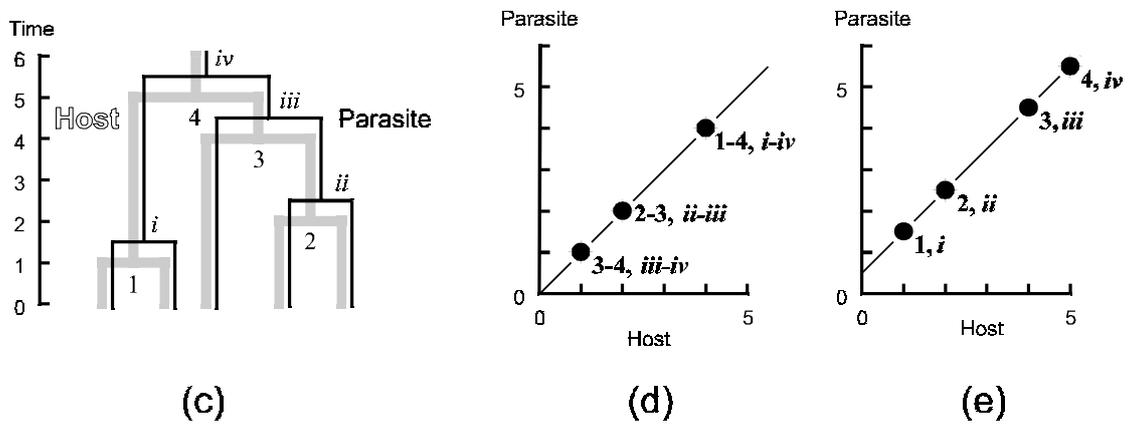


Fig. 6

Temporal Congruence Revisited

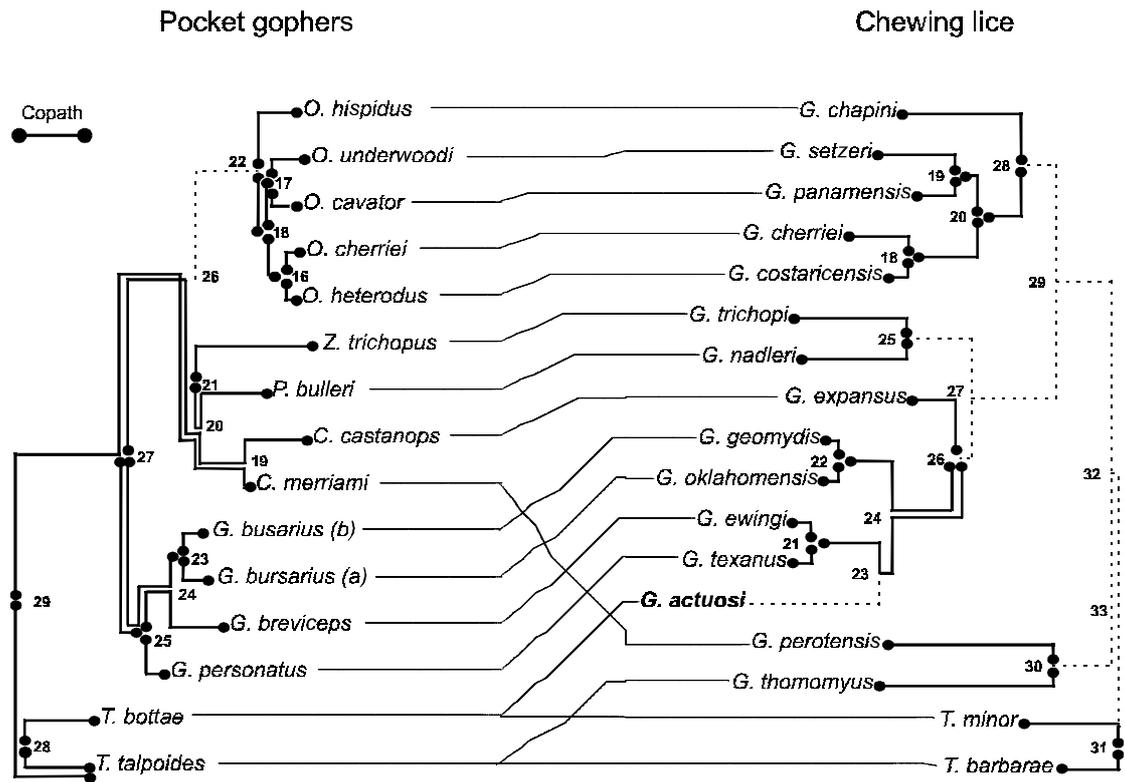


Fig. 7

Temporal Congruence Revisited

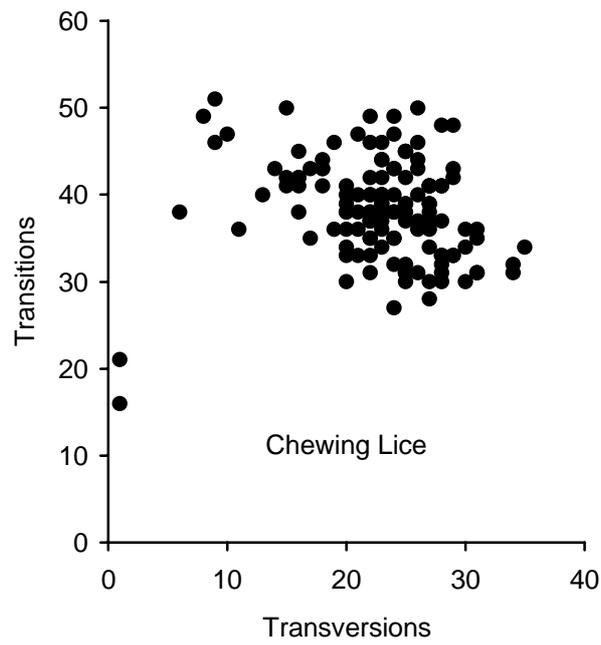
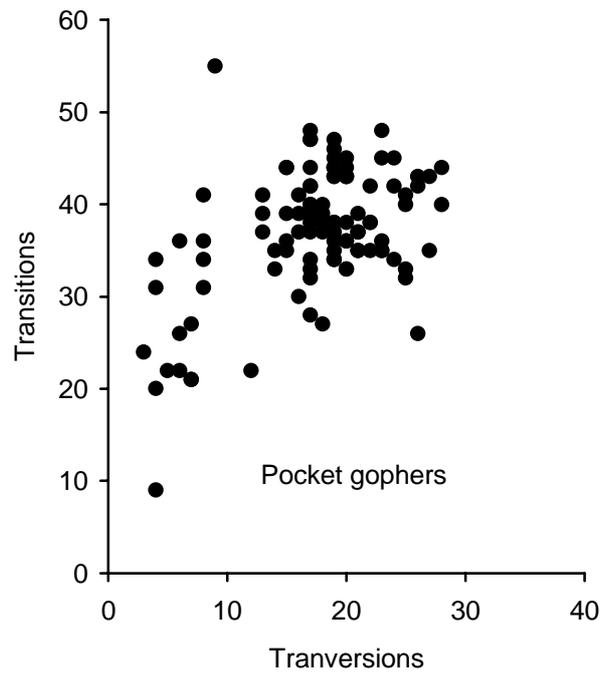


Fig. 8

Temporal Congruence Revisited

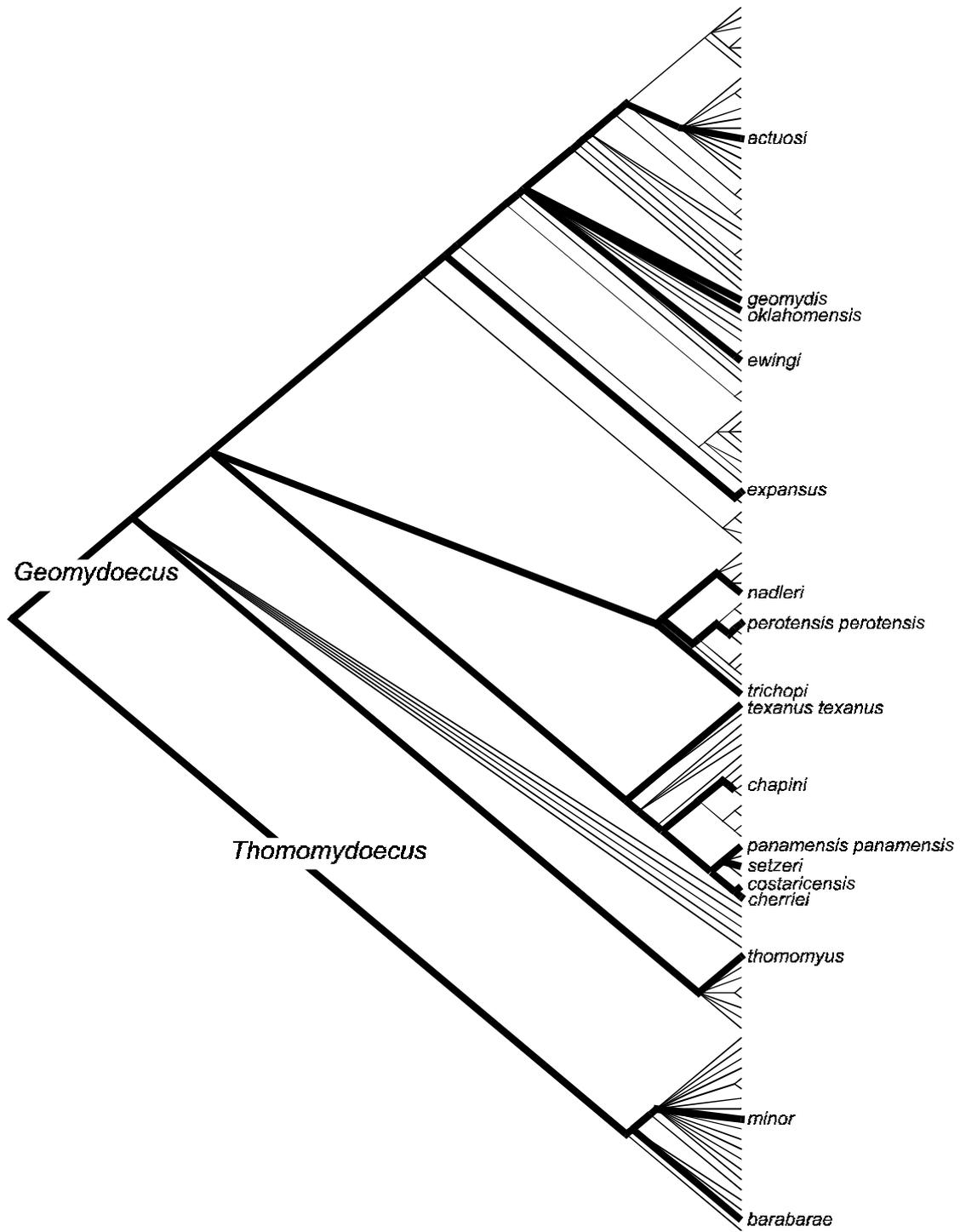


Fig. 9

Temporal Congruence Revisited

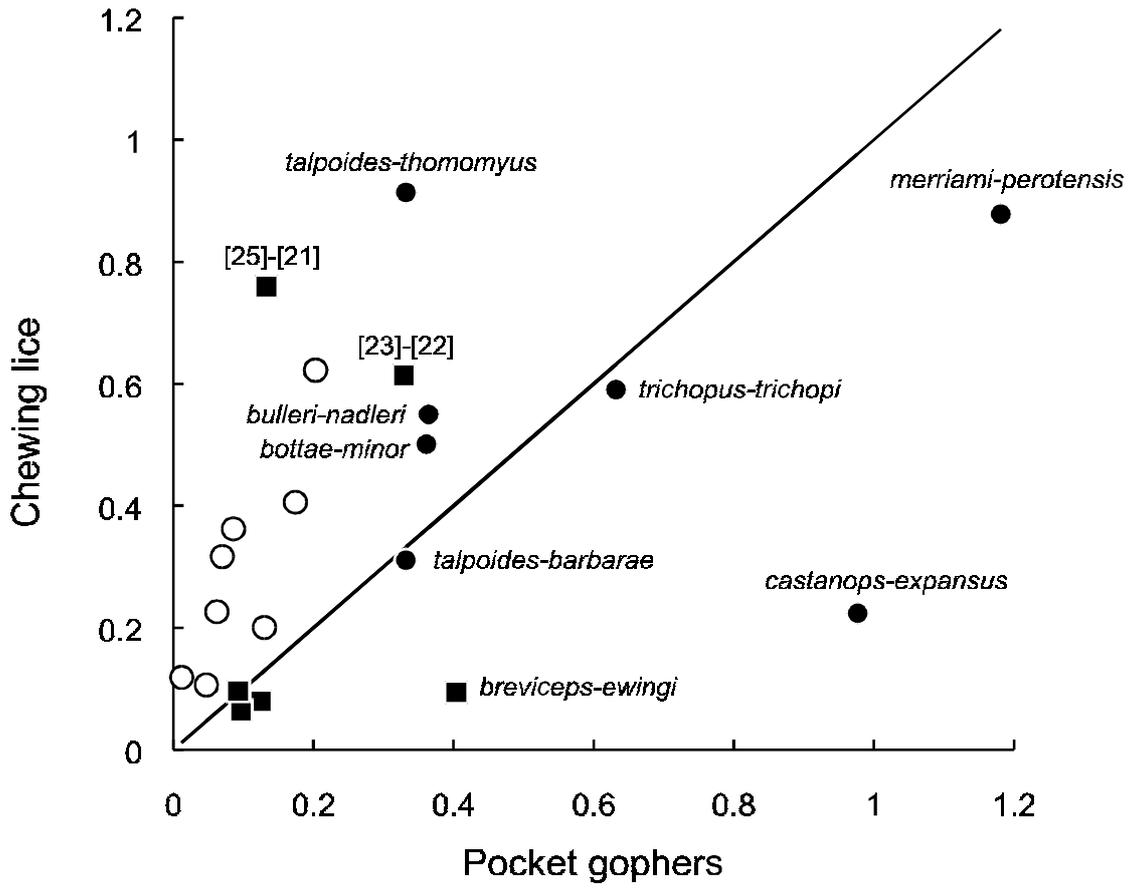


Fig. 10

Temporal Congruence Revisited

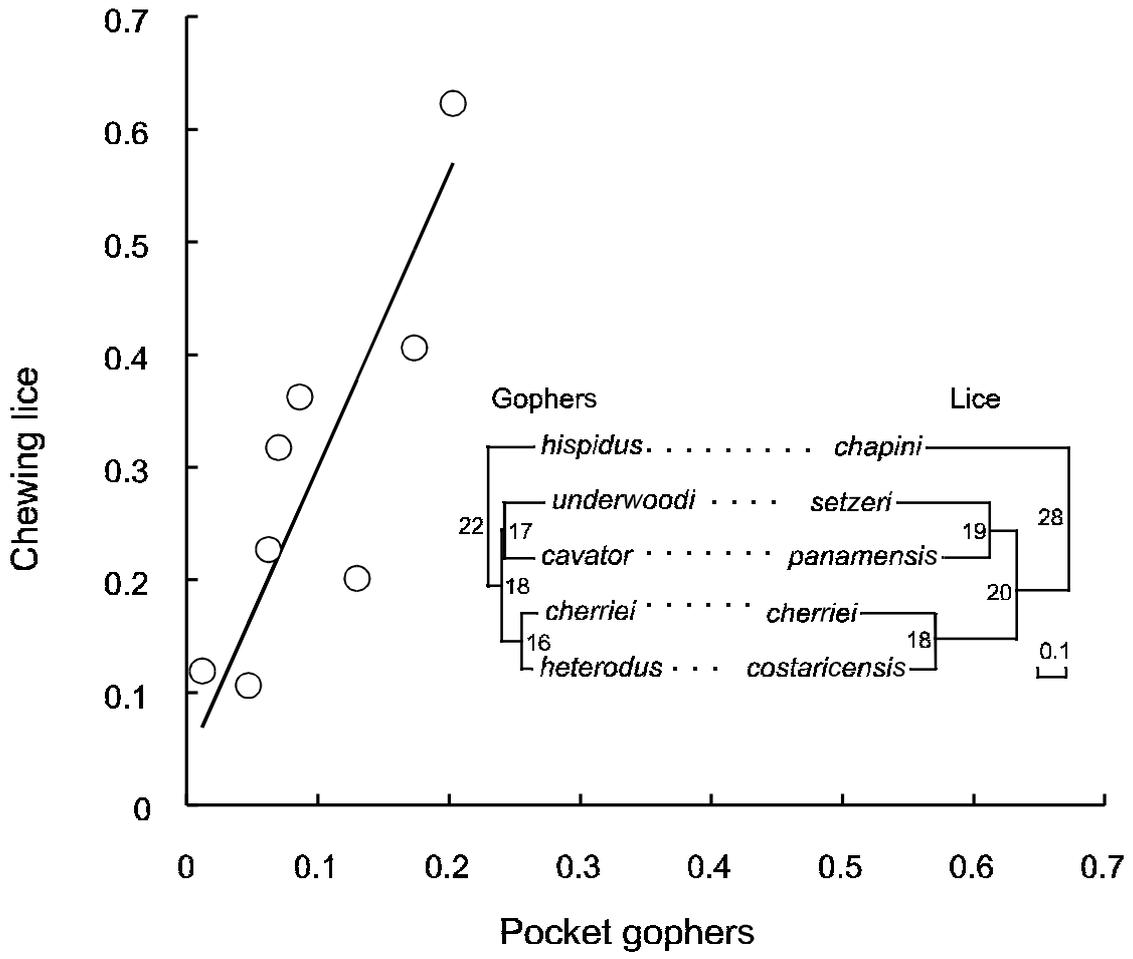


Fig. 11