Combined Data, Bayesian Phylogenetics, and the Origin of the New Zealand Cicada Genera

THOMAS R. BUCKLEY,1,3 PETER ARENSBURGER,2,4 CHRIS SIMON,1,2 AND GEOFFREY K. CHAMBERS1

1Institute for Molecular Systematics, School of Biological Sciences, Victoria University of Wellington, P.O. Box 600, Wellington, New Zealand
2Department of Ecology and Evolutionary Biology, University of Connecticut, Connecticut 06269-3043, USA

Abstract.—We have applied Bayesian and maximum likelihood methods of phylogenetic estimation to data from four mitochondrial genes (COI, COII, 12S, and 16S) and a single nuclear gene (EF1α) from several genera of New Zealand, Australian, and New Caledonian cicada taxa. We specifically focused on the heterogeneity of phylogenetic signal among the different data partitions and the biogeographic origins of the New Zealand cicada fauna. The Bayesian analyses circumvent many of the problems associated with other statistical tests for comparing data partitions. We took an information-theoretic approach to model selection based on the Akaike Information Criterion (AIC). This approach indicated that there was considerable uncertainty in identifying the best-fit model for some of the partitions. Additionally, a large amount of uncertainty was associated with many parameter estimates from the substitution model. However, a sensitivity analysis on the combined dataset indicated that the model selection uncertainty had little effect on estimates of topology because these estimates were largely insensitive to changes in the assumed model. This outcome suggests strong signal in our data. Our analyses support a New Caledonian affiliation of the New Zealand cicada genera *Maoricicada*, *Kikihia*, and *Rhodopsalta* and Australian affinities for the genera *Amphipsalta* and *Notopsalta*. This result was surprising, given that previous cicada biologists suspected a close relationship between *Amphipsalta*, *Notopsalta*, and *Rhodopsalta* based on genitalic characters. Relationships among the closely related genera *Maoricicada*, *Kikihia*, and *Rhodopsalta* were poorly resolved, the mitochondrial data and the EF1α data favoring different arrangements within this clade. [AIC; Bayesian statistics; cicadas; maximum likelihood; model selection; model selection uncertainty; phylogenetics.]

The Bayesian approach to phylogenetics shows considerable promise because it allows simultaneous estimation of the uncertainty associated with any parameter from the phylogenetic model (topology, branch lengths, and substitution model) through the use of the posterior distribution. Because phylogenetic reconstruction is a complex estimation problem, Markov chain Monte Carlo (MCMC) methods are needed to estimate the posterior distribution of the parameters. Bayesian methods have been successfully applied to the estimation of topology and the associated uncertainty (Allison and Wallace, 1994; Rannala and Yang, 1996; Larget and Simon, 1999; Mau et al., 1999; Li et al., 2000; Bollback and Huelsenbeck, 2001; Suchard et al., 2001), dating speciation events (Thorne et al., 1998; Huelsenbeck et al., 2000a; Korber et al., 2000; Wares, 2000; Buckley et al., 2001a), and tests of various evolutionary hypotheses such as evolution of character states (Huelsenbeck et al., 2000b; Huelsenbeck and Bollback, 2001), host–parasite coevolution (Huelsenbeck et al., 2000c), and biogeographic origins (Huelsenbeck and Imennov, 2002). Bayesian methods for quantifying uncertainty have an advantage over some maximum likelihood (ML)–based approaches because they do not require the time-consuming reoptimization of the likelihood function over many replicates, as is the case for bootstrap techniques (Felsenstein, 1985; Goldman et al., 2000). In addition, posterior probabilities are arguably a more intuitive measure of uncertainty than are frequentist *P*-values through being a more direct estimate of confidence in a hypothesis (Gelman et al., 1995).

Here, we focus on the application of Bayesian methods for assessing the congruence of heterogeneous data partitions and on testing biogeographic hypotheses. Testing for incongruence between data partitions is routine in phylogenetics, and numerous tests have been devised to compare data partitions or the trees produced from them (e.g.,
Kishino and Hasegawa, 1989; Rodrigo et al., 1993; Farris et al., 1994; Huelsenbeck and Bull, 1996; Swofford et al., 1996; Shimodaira and Hasegawa, 1999; Waddell et al., 1999, 2000). Parsimony tests of incongruence are more likely to confuse true topological incongruence (different evolutionary history) with systematic error because they lack the inherent ability of model-based methods to accommodate known biases in the data (Huelsenbeck and Bull, 1996; Dolphin et al., 2000; Yoder et al., 2001; Dowton and Austin, 2002). ML tests also have various disadvantages. For example, the nonparametric SH test (Shimodaira and Hasegawa, 1999) requires a priori decisions to be made regarding the number of topologies to be included in the calculation of confidence limits (Goldman et al., 2000). These decisions can drastically affect the size of the $P$-values obtained. Parametric ML tests (Huelsenbeck and Bull, 1996; Swofford et al., 1996; Goldman et al., 2000) are computationally slow and therefore are not useful for many large datasets.

To explore the utility of Bayesian phylogenetic analyses, we used sequence data from four mitochondrial genes and one nuclear gene: 3′ segments of the large (16S) and small (12S) subunit rRNA genes, a portion of subunit I (COI) and all of subunit II (COII) of the cytochrome oxidase gene, and a large region of the nuclear elongation factor $1\alpha$ gene (EF$1\alpha$) and its introns. Sequence data were obtained from 14 New Zealand, New Caledonian, and Australian cicada taxa. The biological motivation for the selection of this taxonomic group was to further understand the biogeographic origins of the New Zealand biota. Recent molecular phylogenetic studies have indicated that many endemic New Zealand taxa have sister species on nearby landmasses. For example, several New Zealand taxa show affinities with New Caledonian taxa (Waters et al., 2000; Boon et al., 2001; Chambers et al., 2001). Phylogenetic analyses of other taxa have indicated an Australian origin, particularly the east coast and Tasmania (Onycophora: Gleeson et al., 1998; Heptalid moths: Brown et al., 1999, 2000; Crayfish: Crandall et al., 1999, 2000; Galaxiid fish: Waters et al., 2000; Microseris: Vijverberg et al., 2000; Ratites: Cooper et al., 2001). Other taxa (Pseudopanax: Mitchell and Wagstaff, 2000; Scrophulariaceae: Wagstaff and Garnock-Jones, 2000; Metrosideros: Wright et al., 2000; parakeets: Boon et al., 2001) appear to have diversified somewhat in New Zealand, subsequently dispersing to other landmasses in the Pacific Ocean. The New Zealand cicada fauna has long been considered an example of oceanic dispersal to New Zealand from neighboring landmasses (Myers, 1929; Dugdale, 1972). However, previous authors were able to only speculate on the geographical origin of the ancestral stock, the number of founding events, and the identity of phylogenetic relatives in the Southwest Pacific (Dugdale, 1972).

**The New Zealand Cicadas**

Five cicada genera (tribe Cicadettini) are found in New Zealand: *Notopsalta* (Dugdale, 1972), *Amphipsalta* (Dugdale and Fleming, 1969), *Rhodopsalta* (Dugdale, 1972), *Maoricicada* (Dugdale, 1972), and *Kikihia* (Dugdale, 1972). All of the genera are endemic except for *Notopsalta* and *Kikihia*, which have one described species on mainland Australia and on Norfolk Island, respectively (Moulds, 1990). However, preliminary (unpublished) data from the COI gene suggest that the Australian *Notopsalta* belong in a different genus. The genus *Amphipsalta* has three species, two of which sing from open vertical surfaces such as tree trunks. One of these species is restricted to the North Island, whereas the other is widely distributed on both islands. Small pockets of a third species of *Amphipsalta* are found today singing from exposed coastal cliffs of the southern coast of the North Island and commonly in river beds, cliffs, and open forests of the northeastern South Island. The genus *Rhodopsalta* also has limited diversity, with only two species found in dry habitats singing from grasses and small shrubs. One species is widely distributed on both islands; the other is restricted to small pockets of dune habitat along the western coast of the North Island. The two most speciose genera are *Maoricicada* and *Kikihia*. The predominantly alpine genus *Maoricicada* contains at least 14 species (Fleming, 1971; Dugdale and Fleming, 1978; Buckley et al., 2001a,b,c). The genus *Kikihia* contains >20 species (Dugdale, 1972; Fleming, 1984), some of which are still undescribed. All of the *Kikihia* species are green or brown and sing from foliage. *Kikihia* species are found in a wide variety of habitats,
including grasslands, scrub, and forests and occupy a wide variety of elevations from sea level to the subalpine region.

Using this combination of mitochondrial markers and a large nuclear gene coupled with ML and Bayesian phylogenetic analyses, we addressed previously unanswered questions relating to New Zealand cicada biogeography. Here, we demonstrate the utility of the Bayesian approach for examining congruence among the different genes and for testing biogeographic hypotheses. We also implement the information-theoretic approach to model selection (Akaike, 1973; Burnham and Anderson, 1998) and assess the relevance of model selection uncertainty (Chatfield, 1995) for analyzing the current data set.

**Materials and Methods**

**Specimen Collection and Laboratory Methods**

Cicada specimens were collected from New Zealand, Australia, and New Caledonia. The New Zealand taxa sampled included *Amphipsalta cingulata*, *Notopsalta sericea*, *Kikihia cauta*, *K. scutellaris*, *Rhodopsalta cruentata*, *R. leptomera*, *Maoricicada cassiope*, and *M. hamiltoni*. Full collection details will be given elsewhere (Arensburger et al., unpubl.). For *Maoricicada* and *Kikihia*, on the basis of extensive intragenic DNA sequencing and phylogenetic analysis (Buckley et al., 2001a,c; Arensburger and Simon, unpubl.) we selected one single basal and one single derived species spanning the root node of that genus to minimize the risk of including long branches in the tree. Luckily, the divergence between the two *Rhodopsalta* species and between *Notopsalta* and *Amphipsalta* are similarly deep. The two New Caledonian taxa sampled were *Pauropsalta johanae* and *Myersalna depicta*, which are among the very few species in the tribe Cicadettini that occur in New Caledonia. The predominantly Australian genus *Cicadetta*, of which we include two species, is highly speciose and is currently being revised (Moulds, unpubl.). The Australian taxa sampled included *Cicadetta celis*, *C. puer*, *Diemeniana frenchi*, and *D. tillyardi* (see Moulds, 1990, for information on the Australian species). Because the tribe Cicadettini has many Australian species in it we narrowed our choice of possible New Zealand cicada relatives on the basis of suggestions from the literature (Dugdale, 1972) and from M. Moulds (pers. comm.). From a subset of 11 species, we narrowed our sample still farther to include taxa least likely to introduce long-branch artifacts (following the strategy of Aguinaldo et al., 1997, and Kim et al., 1999). Branch lengths were estimated based on sequences from the complete COII gene. We selected the genus *Diemeniana* as an outgroup in light of suggestions that it is a close outgroup of New Zealand cicadas (based on morphology; Dugdale, 1972) and is currently placed in the tribe Parmisini, whereas the rest of the species suggested to be close outgroups of New Zealand cicadas are placed in the tribe Cicadettini (Moulds, 1990).

Samples were frozen in the field and stored at about −80°C in the laboratory. At least two individuals from each species were analyzed to control for possible contamination or sample mix-ups. Whole genomic DNA was extracted by the “salting-out” method of Sunnucks and Hales (1996; see also Buckley et al., 2001c, for modifications) or by the CTAB/DTAB method (Gustinich et al., 1991). Portions of the four mitochondrial genes (12S, 16S, COI, and COII) and one nuclear gene (EF1α) were amplified from template DNA by polymerase chain reaction (PCR). PCR products were purified with the QIAquick PCR Purification Kit (Qiagen) or excised from a low-melting-point agarose gel and purified by using the PrepAGene (Bio-Rad) kits. Amplicons were cycle-sequenced by using the Big Dye Terminator Ready Reaction Kit (Perkin-Elmer), and the resulting cycle-sequencing products were separated on an ABI 377 (Automated Biosystems) automated sequencer. Full details of laboratory procedures, including PCR primer sequences and thermal cycling conditions, are given elsewhere (Arensburger et al., unpubl.).

**Sequence Variation and Model Selection**

Alignment of the COI and COII sequences was trivial, given the conserved amino acid sequence reading frame and lack of length variation. The 12S rRNA sequences were aligned by using the conserved motifs and secondary structural elements described by Hickson et al. (1996). The 16S rRNA sequences were aligned according to the secondary structure model of Gutell et al. (1993) and the conserved, insect-specific motifs described in Buckley et al. (2000). Despite the
availability of robust structural information for the rRNA genes, we were forced to exclude some small regions where the alignment was ambiguous. The EF1α exons were also easily aligned because of length and amino acid conservation. The EF1α introns contained indels in some taxa and were aligned by using the program CLUSTAL W (Higgins et al., 1992). Shifts in base composition among taxa were examined by using χ² heterogeneity tests on all sites and on varied sites only.

We took an information-theoretic approach to model selection, based on the Akaike Information Criterion (AIC; Akaike, 1973). We do not favor the more commonly used hypothesis testing approach (Cox, 1962) to phylogenetic model selection for a variety of reasons that have been discussed extensively elsewhere (e.g., Madigan and Raftery, 1994; Kass and Raftery, 1995; Burnham and Anderson, 1998; Forster, 2000, 2001). The AIC is an estimate of the Kullback–Leibler (K-L) distance, which is defined as the information lost when one model is used to approximate another. AIC values give an estimate of the expected distance between an approximating model and the multidimensional, unknown process that has generated the data. The use of normalized Akaike weights (Akaike, 1978) facilitates the identification of groups of models with approximately equal fits to the data (Burnham and Anderson, 1998). Normalized Akaike weights (ωi) for each model (i) are an estimate of the weight of evidence that a given model is the best K-L model in the set of candidate models examined (Burnham and Anderson, 1998). An alternative Bayesian interpretation is that ωi is the conditional probability that model i is the best K-L model given the data (Akaike, 1981).

Model selection uncertainty, or the uncertainty associated with estimating the best K-L model from the set of candidate models examined (Chatfield, 1995), can be quantified by using the normalized Akaike weights—but not if likelihood-ratio tests are used (Burnham and Anderson, 1998). Failure to account for model selection uncertainty will lead to an underestimation of the variance associated with any parameter that is conditional on the model selected (Madigan and Raftery, 1994). AIC values, AIC differences, and normalized Akaike weights were calculated according to Burnham and Anderson (1998) in a Microsoft Excel™ spreadsheet available from the senior author on request. One can also perform model selection by using Bayesian methods (e.g., Schwarz, 1978; Madigan and Raftery, 1994; Kass and Raftery, 1995; Suchard et al., 2001). Like the AIC, the Bayesian methods also enable model selection uncertainty to be quantified and formally incorporated into the inference.

We examined the fit to the data of the same set of candidate models as were utilized by Frati et al. (1997) and Sullivan et al. (1997), including those of Jukes and Cantor (1969; JC69), Kimura (1980; K80), Hasegawa et al. (1985; HKY85), and the general time-reversible model (e.g., Yang, 1994a; GTR). We also accommodated among-site rate variation by using invariable sites (Hasegawa et al., 1985; +I), discrete gamma rates (Yang, 1994b; +Γ), and mixed invariable sites and gamma rates (Gu et al., 1995; I + Γ). We did not investigate site-specific rates models (e.g., Swofford et al., 1996) because these models may give misleading estimates of topology, especially when rates of change vary across the tree (Buckley et al., 2001b; Buckley and Cunningham, in press). We used ML trees estimated under the JC69 model for model fitting. All ML analyses were performed with PAUP*4.0b6–4.0b8 (Swofford, 1998).

Phylogenetic Analyses
ML (Felsenstein, 1981) and Bayesian (Rannala and Yang, 1996; Larget and Simon, 1999; Mau et al., 1999; Li et al., 2000) phylogenetic analyses were performed on each of the five genes separately, on all four mitochondrial genes combined, and finally on all five genes combined (hereafter referred to as the combined dataset). In the analyses in which all of the genes were combined or in which the mitochondrial genes were combined, the model was assumed to be the same for all partitions. Current implementations of ML and Bayesian methods do not allow for sophisticated searches of parameter spaces with partitioned models (e.g., Yang, 1996). For the ML analyses we conducted a heuristic search using TBR branch swapping (start trees were obtained by stepwise addition with 10 random-addition replicates) and the best K-L model indicated by the AIC. ML nodal support was estimated by using the nonparametric
bootstrap (Felsenstein, 1985) with 200 pseudoreplicates and the same heuristic search strategy as described above.

We used MrBayes 1.11 (Huelsenbeck and Ronquist, 2001) to perform the Bayesian phylogenetic analyses. Each Markov chain was started from a random tree and run for $1 \times 10^6$ to $2 \times 10^6$ cycles, sampling every 100th cycle from the chain. We ran four chains simultaneously, three heated and one cold, with the initial 1,000 (10%) cycles discarded as burn-in. To check that stationarity had been reached, we monitored the fluctuating value of the likelihood graphically and repeated each simulation twice. We used the default settings for the priors on the rate matrix (0–100), branch lengths (0–10), gamma shape parameter (0–10), and the proportion of invariant sites (0–1). A Dirichlet distribution was assumed for the base frequency parameters, and an uninformative prior was used for the topology.

We used a Bayesian approach to examine the heterogeneity in phylogenetic signal among the different partitions. For each gene, and for the four mitochondrial genes combined, we estimated the 0.95 posterior intervals (the set of topologies contained in 0.95 of the posterior distribution) for the topology. We estimated the ML topology from a combined analysis of all five genes, assuming the same model for all partitions. We then determined whether the combined analysis topology lay within the 0.95 posterior intervals for each of the five genes and for all of the mitochondrial genes combined. If the combined analysis topology was within the 0.95 posterior intervals for any given partition, then we assumed that that topology could have given rise to the observed data. If the combined analysis topology lay outside the posterior intervals, then one of two possibilities was assumed to be responsible for the incongruence: the particular partition had evolved along another topology, or the assumed substitution model was grossly misspecified. Either possibility has the potential to mislead a combined analysis (Bull et al., 1993). We also compared the phylogenetic signal among the partitions by using Bayes factors (Kass and Raftery, 1995), which measure the strength of evidence of one model relative to another. Kass and Raftery (1995), Suchard et al. (2001), and Huelsenbeck and Immenov (2002) describe how to calculate Bayes factors.

**RESULTS**

**Patterns of Sequence Evolution and Model Selection**

The total dataset after removal of all indels and alignment ambiguous regions consisted of 4,282 sites. The alignment can be obtained from the *Systematic Biology* website (http://www.utexas.edu/depts/systbiol/), and the individual sequences have been deposited in GenBank under the following accession numbers: AF313507, AF313515, AF313517–AF313530, AF313532–AF313537, AF314151–AF314155, AF314157–AF314162, AF247633, AF247626, AF247632, AF426274–AF426287, and AF426438–AF426448. Summary statistics for base frequencies and substitution patterns for each of the genes are given in Table 1. We were not able to reject the hypothesis of base frequency stationarity among the sequences, with the single exception of the varied sites from the COI gene (Table 1). However, our unpublished data suggest that this bias has not had a strong influence on the phylogenetic analyses (Arensburger et al., unpubl.).

Identifying the best K-L model for the 12S, 16S, COI, and combined dataset involved a high level of model selection uncertainty. For example, for the 12S gene, the best K-L model was the GTR + Γ model, which had an $\omega_i$ value of only 0.563 (Table 2), whereas

---

**TABLE 1.** Summary of base frequency tests and substitution patterns for the EF1α, 12S rRNA, 16S rRNA, COI, and COII genes and for the combined dataset. The base frequency test P-values were calculated by using the $\chi^2$ test in PAUP*4.0. MP refers to maximum parsimony sites.

<table>
<thead>
<tr>
<th></th>
<th>EF1α</th>
<th>12S</th>
<th>16S</th>
<th>COI</th>
<th>COII</th>
<th>All sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of sites</td>
<td>2,033</td>
<td>371</td>
<td>432</td>
<td>753</td>
<td>693</td>
<td>4,282</td>
</tr>
<tr>
<td>Varied sites</td>
<td>432</td>
<td>108</td>
<td>111</td>
<td>243</td>
<td>256</td>
<td>1,150</td>
</tr>
<tr>
<td>MP sites</td>
<td>119</td>
<td>59</td>
<td>54</td>
<td>143</td>
<td>149</td>
<td>524</td>
</tr>
<tr>
<td>Base frequency test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All sites</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>Varied sites</td>
<td>1.00</td>
<td>0.97</td>
<td>0.99</td>
<td>0.01</td>
<td>0.24</td>
<td>0.47</td>
</tr>
<tr>
<td>Model</td>
<td>12S $\Delta_1$</td>
<td>16S $\Delta_1$</td>
<td>COI $\Delta_1$</td>
<td>COII $\Delta_1$</td>
<td>All mitochondrial $\Delta_1$</td>
<td>EF1ar $\Delta_1$</td>
</tr>
<tr>
<td>----------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>JC69</td>
<td>390.266</td>
<td>427.058</td>
<td>983.867</td>
<td>935.683</td>
<td>2,771.056</td>
<td>384.418</td>
</tr>
<tr>
<td>K80</td>
<td>367.730</td>
<td>413.842</td>
<td>773.739</td>
<td>668.181</td>
<td>2,281.653</td>
<td>319.382</td>
</tr>
<tr>
<td>HKY</td>
<td>202.787</td>
<td>211.344</td>
<td>627.324</td>
<td>469.319</td>
<td>1,590.927</td>
<td>113.983</td>
</tr>
<tr>
<td>GTR</td>
<td>135.598</td>
<td>169.334</td>
<td>474.660</td>
<td>394.043</td>
<td>1,230.373</td>
<td>107.257</td>
</tr>
<tr>
<td>JC69 + Γ</td>
<td>269.855</td>
<td>276.787</td>
<td>591.284</td>
<td>654.034</td>
<td>1,773.834</td>
<td>276.822</td>
</tr>
<tr>
<td>K80 + Γ</td>
<td>246.227</td>
<td>262.675</td>
<td>371.465</td>
<td>370.784</td>
<td>1,259.766</td>
<td>209.580</td>
</tr>
<tr>
<td>HKY85 + Γ</td>
<td>22.764</td>
<td>7.942</td>
<td>28.429</td>
<td>9.409</td>
<td>83.027</td>
<td>5.401</td>
</tr>
<tr>
<td>GTR + Γ</td>
<td>0.563</td>
<td>0.717</td>
<td>10.582</td>
<td>3.552</td>
<td>9.856</td>
<td>0.936</td>
</tr>
<tr>
<td>JC69 + I</td>
<td>278.059</td>
<td>286.043</td>
<td>607.731</td>
<td>678.779</td>
<td>1,840.162</td>
<td>291.336</td>
</tr>
<tr>
<td>K80 + I</td>
<td>255.101</td>
<td>272.763</td>
<td>392.004</td>
<td>400.746</td>
<td>1,335.681</td>
<td>224.455</td>
</tr>
<tr>
<td>HKY85 + I</td>
<td>51.065</td>
<td>31.002</td>
<td>49.590</td>
<td>52.780</td>
<td>209.385</td>
<td>19.265</td>
</tr>
<tr>
<td>GTR + I</td>
<td>16.074</td>
<td>20.164</td>
<td>31.571</td>
<td>48.619</td>
<td>123.526</td>
<td>14.267</td>
</tr>
<tr>
<td>JC69 + I + Γ</td>
<td>271.484</td>
<td>278.787</td>
<td>590.398</td>
<td>655.073</td>
<td>1,773.947</td>
<td>278.821</td>
</tr>
<tr>
<td>K80 + I + Γ</td>
<td>247.801</td>
<td>264.675</td>
<td>370.871</td>
<td>371.971</td>
<td>1,259.461</td>
<td>211.577</td>
</tr>
<tr>
<td>HKY85 + I + Γ</td>
<td>23.598</td>
<td>9.942</td>
<td>7.339</td>
<td>2.296</td>
<td>45.590</td>
<td>—</td>
</tr>
<tr>
<td>GTR + I + Γ</td>
<td>0.509</td>
<td>0.437</td>
<td>1.996</td>
<td>0.669</td>
<td>0.993</td>
<td>—</td>
</tr>
</tbody>
</table>
the GTR + I + Γ model had an \( \omega_i \) value of 0.437. This result indicates that the information in the 12S dataset is insufficient to be able to distinguish between the two distributions of among-site rate variation (Γ vs. I + Γ) with any certainty. Similarly, the normalized Akaikeweights from the combined dataset for the GTR + Γ and GTR + I + Γ models were 0.307 and 0.693, respectively—again suggesting model selection uncertainty.

The AIC indicated that the GTR + I + Γ model was the best K-L model for the mitochondrial genes and the combined dataset, whereas the GTR + Γ model was the best K-L model for the EF1α gene. The \( \omega_i \) values (Table 2) of the best K-L models were very high for the mitochondrial data (0.993) and the EF1α data (0.936), indicating a very high level of certainty that the best K-L models were identified. For all character partitions, the equal rates and invariable sites models were identified as having very poor fits to the data.

The 0.95 posterior intervals on the model parameter estimates (Table 3) indicate considerable uncertainty in these estimates, even for the large combined datasets. For example, the most commonly occurring substitution type in the mitochondrial dataset was the A→G transition type, for which the 0.95 posterior intervals ranged from 44.808 to 99.360 (Table 3). The posterior intervals for the \( \alpha \) shape parameter under the mixed I + Γ models were also very wide, in agreement with the normalized Akaikeweights. In combination, these analyses suggest considerable uncertainty in identifying the distribution of among-site rate variation.

**Phylogenetic Relationships Among Cicada Genera and Incongruence Among Partitions**

If tests of topology indicate that different partitions do not share the same evolutionary history, then combining such datasets may be unwise (Bull et al., 1993). The Bayesian approach we implemented indicates that the ML topology obtained from the combined analysis (see below) lies within the 0.95 posterior intervals for each of the five genes and for all of the mitochondrial genes combined (Table 4). This is consistent with the hypothesis that each of the partitions we examined has evolved along the same underlying topology, such that a combined analysis is warranted (Bull et al., 1993). The 0.95 posterior intervals for the four mitochondrial genes analyzed separately contained very large numbers of unique topologies (3,326–6,480) relative to the EF1α gene (8 topologies), the mitochondrial dataset (5 topologies), and the combined dataset

<table>
<thead>
<tr>
<th>Data partition</th>
<th>12S</th>
<th>16S</th>
<th>COI</th>
<th>COII</th>
<th>Mito.</th>
<th>EF1α</th>
<th>All sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pr</td>
<td>&lt;0.0001</td>
<td>0.005</td>
<td>0.020</td>
<td>&lt;0.0001</td>
<td>0.524</td>
<td>0.017</td>
<td>0.246</td>
</tr>
<tr>
<td>Cumulative posterior intervals</td>
<td>0.908</td>
<td>0.012</td>
<td>0.228</td>
<td>0.778</td>
<td>0.524</td>
<td>0.942</td>
<td>0.246</td>
</tr>
<tr>
<td>No. of trees in 0.95 posterior intervals</td>
<td>4,494</td>
<td>3,326</td>
<td>3,840</td>
<td>6,480</td>
<td>5</td>
<td>8</td>
<td>5</td>
</tr>
</tbody>
</table>
FIGURE 1. ML trees estimated from the combined mitochondrial data (a), EF1α gene (b), and all sites combined (c). Posterior probabilities followed by nonparametric bootstrap proportions are given adjacent to each node. ML branch lengths are drawn proportional to the number of substitutions per site, as indicated by the scale bars.
These observations reflect the dramatic increase in resolving power when the individual mitochondrial genes are combined.

The ML trees for the EF1α gene and for the combined dataset were identical to the topologies with the greatest posterior probabilities. The ML tree obtained from the combined mitochondrial data differed from the topology with the greatest posterior probability (Pr value) in the arrangement of the two New Caledonian species, which were sister taxa in the Bayesian analysis (data not shown). Phylogenetic analysis of the combined mitochondrial data favored a sister group relationship between Maoricicada and Rhodopsalta (Pr = 0.932, BP = 58%), whereas the EF1α gene favored a clade containing Kikihia and Rhodopsalta (Pr = 0.938, BP = 72%). When all of the sites were combined, the Maoricicada + Rhodopsalta clade was optimal (Pr = 0.487, BP = 48%); however, the alternative Kikihia + Rhodopsalta grouping received almost the same level of support (Pr = 0.508, BP = 51%). Support for a clade containing Kikihia and Maoricicada was virtually nonexistent in analyses of the EF1α data (Pr = 0.003, BP < 5%), the mitochondrial data (Pr = 0, BP < 1%), and all sites combined (Pr = 0, BP < 5%).

The other region of incongruence among the three datasets was the arrangement of the two New Caledonian species and the Kikihia + Maoricicada + Rhodopsalta clade. Both the EF1α gene and the mitochondrial data favored a clade containing Myersalna de picta and the New Zealand genera Kikihia, Maoricicada, and Rhodopsalta. However, support for this arrangement was weak (Fig. 1). In the combined dataset the New Caledonian species were monophyletic; however, support was again weak (Pr = 0.507, BP = 40%). The ML analyses showed that the internal branches that resolve this New Caledonian clade are all extremely short.

We plotted the posterior probabilities of all of the topologies that were included in the posterior intervals for the combined mitochondrial data, the EF1α gene, and all sites combined (Fig. 2). These results show that the phylogenetic signal, although not strongly conflicting, is not strictly additive when the mitochondrial genes and the EF1α gene are combined. Comparing the Bayes factors calculated by comparing the optimal topologies from the different partitions also supports these results. The Bayes factor from the EF1α data calculated by comparing the optimal EF1α topology and the combined analysis topology is 24.55, which indicates strong support for the EF1α topology relative to the topology obtained from the mitochondrial genes and the combined analysis. However, when the EF1α topology and the combined analysis topology are compared by using the mitochondrial dataset, the Bayes factor in support of the combined analysis topology is 58.70, indicating strong support. When the two topologies are compared by using all the combined dataset, the Bayes factor is 1.97, which suggests that the phylogenetic signal is not strictly additive and that the combined dataset is unable to differentiate between these two hypotheses.

We tested the hypothesis that the five New Zealand taxa form a monophyletic group and are thus descended from a single migration event. We were able to exclude the hypothesis of monophyly because the posterior probability of this node was extremely low (<0.05) for the EF1α data, the combined mitochondrial data, and the combined dataset. Bootstrap support for the monophyly of the five New Zealand genera was also nonexistent. These results are consistent with the alternative hypothesis that the New Zealand cicada fauna is polyphyletic, resulting from at least two migrations.

Because the AIC results indicated that those data were unable to distinguish between the GTR + Γ model and the GTR + I + Γ mode, we conducted a sensitivity analysis on the combined dataset, examining the effects of shifts in substitution model assumptions on the posterior distribution. The posterior probabilities under both models were very similar. For example, the same topology
was optimal under both models, and the largest difference in posterior probability for a single node was 0.032 for the node joining the two New Caledonian species.

**DISCUSSION**

*Bayesian Methods for Examining Heterogeneous Data*

The Bayesian analyses indicated that the optimal topology from the combined analysis lay within the 0.95 posterior intervals for each of the individual partitions (Table 4). Because our expectation is that each of the mitochondrial genes evolved along the same underlying topology, a finding that the combined analysis topology lay outside these intervals would indicate severe model misspecification. Combining the four mitochondrial genes also led to a substantial increase in phylogenetic resolution, such that the posterior intervals shrunk from several thousand topologies to only five topologies.

The phylogenetic analysis of the EF1α gene and the combined mitochondrial dataset suggested general agreement between these datasets, except with regard to the relationships within the *Kikihia, Maoricicada*, and *Rhodopsalta* clade. Based on the Bayesian analysis, the EF1α gene strongly supported a *Kikihia + Rhodopsalta* clade, whereas the mitochondrial dataset strongly supported a *Maoricicada + Rhodopsalta* grouping. Interestingly, with respect to the *Kikihia, Maoricicada*, and *Rhodopsalta* clade, the signal within both of the partitions does not appear to be strictly additive when the data are combined (Fig. 2). The posterior probabilities of the two conflicting nodes within this clade for the mitochondrial data and the EF1α gene are 0.932 and 0.938, respectively. When all of the partitions are combined, both conflicting nodes have posterior probabilities of ~0.5. Lack of additive signal between the two partitions is not necessarily indicative of different evolutionary histories, although we cannot entirely rule this out, given the short internal branches. However, if the assumed model was grossly misspecified for one of the partitions, causing an incorrect topology to be preferred, then we might expect to observe mixed support for the two conflicting nodes when the data are combined.

The ML bootstrap values display the same pattern as the posterior probabilities, although the values are less extreme. Despite the fundamental statistical differences between the two measures of support, the large difference in statistical support between the Bayesian and ML methods for these nodes is worth further investigation. We suspect that the inherent, conservative nature of the non-parametric bootstrap (Hillis and Bull, 1993) is reflected in the more moderate levels of support for the conflicting nodes. Additionally, Bayesian methods may be more susceptible to model misspecification than some non-parametric ML tests (Shimodaira, 2001). If this is the case with some of our data, then we might expect to observe inflated Bayes factors and posterior probabilities for individual nodes. Interestingly, other nodes (e.g., *Rhodopsalta + Maoricicada* in the combined dataset) received very similar (low) support from both the ML and the Bayesian analyses (Fig. 1c).

If the EF1α gene and the mitochondrial partition had different evolutionary histories, then we would expect the optimal topology from each partition to lie outside the 0.95 posterior intervals, assuming the data contained enough signal. Although both optimal topologies had very low posterior probabilities from the opposing partitions, lying just within the 0.95 posterior intervals (Table 4), sampling error exacerbated by model misspecification cannot be excluded. Because we are combining several heterogeneous datasets with quite different selective constraints (i.e., protein-coding and rRNA genes), we believe model misspecification is a strong possibility. If we were able to implement the Bayesian analyses in conjunction with partitioned likelihood models (e.g., Yang, 1996), then we would expect to observe changes in the posterior distribution of topologies, although what effects this would have are difficult to anticipate.

An area of recent activity in the phylogenetic methodology literature has been tests for comparing data partitions. Some tests (e.g., the partition homogeneity or incongruence length difference test; Farris et al., 1994) ask whether combining data significantly increases the homoplasy present in a phylogenetic analysis. Others (e.g., the likelihood heterogeneity test; Huelsenbeck and Bull, 1996) ask whether the same phylogenetic tree underlies two data partitions. Untangling the effects of topological incongruence and systematic error in a parsimony framework (e.g., Baker et al., 2001; Sota and Vogler, 2001)
is problematic because accounting for potentially misleading features of the substitution process is much more difficult than in a model-based framework. Probabilistic methods have the advantage of being able to minimize the impact of systematic error through the use of flexible models. Recent empirical studies have revealed problems with using the partition homogeneity test (Farris et al., 1994) as a guide to dataset combinability (e.g., Dolphin et al., 2000; Yoder et al., 2001; Dowton and Austin, 2002). Additionally, measures such as partition Bremer support values (Baker et al., 1998) are difficult to interpret statistically (DeBry, 2001), although they are useful for estimating the number of character-state changes that support various nodes from different partitions.

The approach we have taken is conceptually similar to the ML tests of Huelsenbeck and Bull (1996) and Waddell et al. (2000): All are an attempt to estimate the uncertainty associated with a topology from one partition and determine whether a topology from another partition lies within that region of uncertainty. However, the Bayesian method we have used here has two advantages over ML-based methods. First, the Bayesian approach has a computational advantage over parametric bootstrap methods (e.g., Huelsenbeck and Bull, 1996; Swofford et al., 1996; Goldman et al., 2000), which require reoptimizing the likelihood over many replicates (although see Waddell et al., 2000). Second, implementing the nonparametric SH test (Shimodaira and Hasegawa, 1999; Buckley et al., 2001c) can be problematic because how many, and exactly which, topologies to include in the set of candidate topologies to be tested is not clear. Including too many topologies increases the conservativeness of the test and reduces its power to detect any incongruence. Including too few topologies may lead to Type I errors and the unwarranted rejection of congruence. This second problem is circumvented by the Bayesian approach because no a priori decisions need be made regarding what topologies to include when estimating the posterior distribution.

The Relevance of Model Selection Uncertainty

We also assessed the relevance of model selection uncertainty (Chatfield, 1995) on the phylogenetic analysis of the combined dataset. The AIC analyses indicated that for the combined dataset we were unable to identify the best K-L model with any confidence. In the case of the individual mitochondrial genes, there was also considerable model selection uncertainty—again, between the GTR + Γ and GTR + I + Γ models. The observed uncertainty in model selection probably reflects the difficulty in determining whether constant sites are actually slowly evolving (i.e., as described in the gamma distribution) or are truly invariable (i.e., as in an invariable sites model), when only a few sequences are available (Sullivan et al., 1999). We expect the addition of further taxa would increase our confidence that one of the two models is the best K-L model. The addition of more models, with improved fit to the data, would also alter the Akaike weights for the individual models we examined here.

Our failure to find a model of evolution that increases the congruence among the mitochondrial genes and the EF1α gene is not surprising. Finding a single model to fit the evolution of a large rRNA gene is difficult because RNA molecules have alternating regions of various lengths and probabilities of substitution (Simon et al., 1994; Kjer, 1995; Hickson et al., 1996; Buckley et al., 2000). Combining rRNA and protein-coding genes is even more problematic. Of the models that explicitly allow for spatial variation in the pattern of nucleotide substitution (Yang, 1996), none has yet been implemented in conjunction with sophisticated tree-searching algorithms or with Bayesian MCMC methods.

The presence of model selection uncertainty indicates that it may be wise to base further estimation on both the Γ model and the I + Γ model (Madigan and Raftery, 1994; Buckland et al., 1997). For the combined dataset, however, the estimates of posterior probabilities were very similar between the two models. This result suggests that the inability to identify the best K-L model with any confidence is not a serious problem for these data. Because our knowledge regarding the effects of small shifts in model structure on phylogenetic reconstruction is still limited, we recommend performing sensitivity analyses across reasonable models (especially any plausible among-site rate variation models) to check that the conclusions drawn are stable (Buckley et al., 2001; Buckley and Cunningham, in press).
More advanced methods exist for formally accommodating model selection uncertainty, such as Bayesian model averaging (e.g., Madigan and Raftery, 1994; Hoeting et al., 1999) and extensions of the AIC framework (Buckland et al., 1997; Burnham and Anderson, 1998). Under Bayesian model averaging, the posterior probability of a parameter value is a weighted average of the posterior probabilities estimated under various models. This can be achieved by using MCMC algorithms that move simultaneously in parameter and model space (e.g., Green, 1995). The AIC-based methods for accommodating model selection uncertainty operate by utilizing the Akaike weights to obtain weighted estimates of parameter values. The AIC-based methods have the advantage of being computationally fast relative to MCMC methods. Given that model averaging methods are still being developed, and in some cases are technically demanding, we believe that sensitivity analyses offer a convenient alternative (Chatfield, 1995).

The high level of uncertainty associated with the parameter estimates of the substitution model (Table 3) is somewhat surprising given the long sequence length of the combined dataset (4,282 sites). As with the AIC analyses, we would expect these posterior intervals to narrow with the addition of further taxa (see Excoffier and Yang, 1999; Sullivan et al., 1999). The high level of uncertainty associated with the parameter estimates suggests a further advantage for the hierarchical Bayesian approach, which, unlike most ML analyses, does not require the parameter values to be fixed during phylogenetic estimation.

Biogeographic and Evolutionary Implications

Our results strongly support the hypothesis that the New Zealand cicada fauna is derived from two independent migrations. Importantly, these results add to the growing body of data that indicate strong biogeographical links between New Zealand and New Caledonia (Waters et al., 2000; Boon et al., 2001; Chambers et al., 2001; Lee et al., 2001). Possibly the divergences among the three faunas are Gondwanan, or alternatively, of more recent oceanic dispersal. The three genera Kikihia, Maoricicada, and Rhodopsalta may have dispersed to New Zealand along island arcs from New Caledonia during the Oligocene (Herzer et al., 1997; Herzer, 1998). Testing these hypotheses will involve the application of molecular dating approaches (e.g., Thorne et al., 1998). Such studies are currently in progress (Arensburger et al., unpubl.; Simon et al., unpubl.).

The general position of Rhodopsalta in the topologies we have obtained has implications for the evolution of cicada genitalia and song structure. The genera Rhodopsalta, Amphipsalta, and Notopsalta have a trifid aedeagus (male intromittent organ), and small pygophore (abdominal segment IX that houses the aedeagus). For this reason, Dugdale (pers. comm.) believed that Amphipsalta, Notopsalta, and Rhodopsalta formed a monophyletic group (hence the names all ending in “psalta”). The pygophore is much larger in Kikihia and in Maoricicada (Dugdale, 1972), and the aedeagus is less sclerotized. Unlike all of the remaining species in this study, both Rhodopsalta and Maoricicada have lost the ability to produce an alarm call. A single loss of this trait would be consistent with the mitochondrial dataset. Fleming (1975) noted similarities in song structure between Maoricicada and Kikihia, a grouping not supported by any of the partitions in our data, and in conflict with a single loss of the alarm-call character. Detailed studies of the evolution of song components among the >40 species of New Zealand cicadas are currently in progress.

Conclusions

We have illustrated how Bayesian methods for quantifying uncertainty can be adapted for comparing the phylogenetic signal between different data partitions through the use of posterior probabilities. We believe that this general approach for examining incongruence has several advantages over both parsimony-based and ML methods. These advantages include computational efficiency and the fact that, unlike the SH test, MCMC methods remove the requirement to specify the total set of topologies a priori to the calculation of confidence levels. The data presented here also illustrate how the combination of heterogeneous datasets can obscure underlying incongruence and does not necessarily lead to more strongly supported estimates of phylogeny.
ACKNOWLEDGMENTS

We are grateful to the following people for help and assistance with field work, locality information, or the identification of specimens: Michel Bouard, Steve Chiswell, David Emery, and Max Moulds. T.R.B. thanks Liz MacAvoy, Wee Ming Boon, Smita Apte, Lesley Millicich, Darren Day, and Paul Sunnucks for advice/assistance with laboratory work. Jonathan Gardner, David Penny, Hirohisa Kishino, Richard Olmstead, Ted Schultz, Jeff Thorne, and an anonymous reviewer provided comments on the manuscript. John Huelsenbeck provided an unpublished manuscript and advice on analyses. The Department of Conservation (Te Papa Atawhai) and Tasmania National Parks provided collecting permits and information on access to field sites. Funding was provided by The National Geographic Society, Victoria University of Wellington, the National Science Foundation, the Fullbright Foundation, and the University of Connecticut.

REFERENCES


