Systematic Biology Advance Access published October 6, 2011

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Running head: MULTILOCUS PHYLOGENY OF FAIRY-WRENS

A Species Tree for the Australo-Papuan Fairy-wrens and Allies (Aves: Maluridae)

June Y. Lee[1,3], Leo Joseph[2], and Scott V. Edwards[1,3]

[1]Museum of Comparative Zoology and Department of Organismic & Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138

[2] Australian National Wildlife Collection, CSIRO Ecosystem Sciences, GPO Box 284, Canberra, ACT 2601, Australia

[3]These authors contributed equally to this manuscript.

Author for Correspondence:
Scott V. Edwards
Museum of Comparative Zoology
Harvard University
26 Oxford Street
Cambridge MA 02138
phone: 617-496-9389
fax: 617-495-5667

© The Author(s) 2011. Published by Oxford University Press, on behalf of the Society of Systematic Biologists. All rights reserved.
For Permissions, please email: journals.permissions@oup.com

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

Abstract. -We explored the efficacy of species tree methods at the family level in birds, using the Australo-Papuan Fairy-wrens (Passeriformes: Maluridae) as a model system. Fairy-wrens of the genus *Malurus* are known for high intensities of sexual selection, resulting in some cases in rapid speciation. This history suggests that incomplete lineage sorting (ILS) of neutrally evolving loci could be substantial, a situation that could compromise traditional methods of combining loci in phylogenetic analysis. Using eighteen molecular markers (5 anonymous loci, 7 exons, 5 introns and one mtDNA locus), we show that gene tree monophyly across species could be rejected for 16 out of 18 loci, suggesting substantial ILS at the family level in these birds. Using the software Concaterpillar, we also detect three statistically distinct clusters of gene trees among the 18 loci. Despite substantial variation in gene trees, species trees constructed using four different species tree estimation methods (BEST, BUCKy, and STAR) were generally well-supported and similar to each other and to the concatenation tree, with a few mild discordances at nodes that could be explained by rapid and recent speciation events. By contrast, minimizing deep coalescences (MDC) produced a species tree that was topologically more divergent from those of the other methods as measured by multidimensional scaling of trees. Additionally, gene and species trees were topologically more similar in the BEST analysis, presumably because of the species tree prior employed in BEST which appropriately assumes that gene trees are correlated with each other and with the species tree. Among the 18 loci we also discovered 102 independent indel markers, which also proved phylogenetically informative, primarily among genera, and displayed a ~4-fold bias towards deletions. As suggested in earlier work, the grasswrens (*Amytornis*) are sister to the rest of the family, and the emu-wrens (*Stipiturus*) are sister to fairy-wrens (*Malurus, Clytomias*). Our study shows that ILS is common at the family level in birds yet, despite this, species tree methods converge on broadly similar results for this family.

Keyword: Incomplete lineage sorting, indel, Maluridae, species tree, passerine, biogeography

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

In the field of phylogenetics, single-locus approaches using mitochondrial DNA (mtDNA) have proved powerful because of mtDNA's advantages over other genetic markers: relatively small population size, high mutation rate, putative lack of recombination and ease of access to due to its high copy number and the availability of primer sequences and whole genomes (Avise 2000; Zink and Barrowclough 2008). These advantages provide relatively better resolution of phylogenetic relationships compared to the other genetic markers on a per-locus basis. However, even mtDNA is susceptible to the stochasticity that is frequently observed in gene genealogies, resulting in gene trees that may not faithfully track the history of speciation events (Avise 2000). One of the most common stochastic processes is incomplete lineage sorting (ILS), which occurs when genetic drift has not had enough time to bring individual gene loci to fixation and/or reciprocal monophyly before subsequent divergence. If this happens, the genealogical histories of individual gene loci may appear misleading about relationships among species (Funk and Omland 2003; Maddison and Knowles 2006). Coalescent theory indicates that this is more likely to happen when the population size of ancestral branches is large relative to the divergence time of two daughter species, such as might occur in recent divergences or rapid radiations. One way to overcome or reduce the effects of this stochasticity is to increase the number of loci studied (Knowles 2009; Edwards 2009).

Sequence-based genetic markers can be divided into several different categories, including mtDNA, introns, exons and anonymous loci (Brito and Edwards 2009). These types of markers have had various levels of success in avian phylogenetic studies, with exons sometimes showing more limited powers of resolution than introns (Chojnowski et al. 2008) and anonymous loci (Karl and Avise 1993) showing substantial amounts of variation at several phylogenetic levels (Jennings and Edwards 2005; Thomson et al. 2008). Additionally, anonymous markers possess

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

abundant variation, exceeding that found in introns, at least within species (Lee and Edwards 2008). However, anonymous loci can sometimes be challenging to characterize in all species of a given clade, given the lack of conserved flanking regions such as possessed by introns (Thomson et al. 2008). An equally challenging aspect of multilocus phylogenetic analysis is how to combine data from different loci. The most commonly used approach is to analyze concatenated data sets, which assume that all genes have a single tree congruent with the species tree. However, the common signal extracted from large, concatenated data sets does not always overcome the conflicting signals that can be produced at individual loci (Kubatko and Degnan 2007). An emerging solution is to incorporate models of stochastic mutation with gene coalescence directly into the estimation of species trees, although this task is analytically challenging (Maddison 1997; Felsenstein 2004; Liu et al. 2009). Recently, methods that estimate a species tree directly by incorporating heterogeneity in gene trees have been introduced (Maddison and Knowles 2006; Liu and Pearl 2007; Liu 2009; reviewed in Liu et al. 2009). Although still early in their development, these methods suggest that with a reasonable number of loci and individuals, it will be possible to infer lineage relationships despite ILS (Maddison and Knowles 2006; Edwards et al. 2007).

Fairy-wrens and Allies

The Australo-Papuan avian family Maluridae comprises 26 species and is distributed throughout Australia and New Guinea (Schodde 1982; Rowley and Russell 1997). Of the family's five genera, *Malurus* has been most studied to date because of its extraordinary behavioral ecology, which includes extremely frequent extra-pair copulation, high sexual dimorphism, cooperative breeding, and intense sperm competition (Brooker et al. 1990; Mulder and Cockburn 1993; Rowe et al. 2008; Webster et al. 2008). In contrast, the other genera have

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

not been as intensively studied mostly due to their relatively low population densities and often more remote ranges (Rowe and Pruett-Jones 2008). The genera *Clytomias* and *Sipodotus* are both monotypic and endemic to New Guinea and outlying Indonesian islands, whereas the emu-wrens (*Stipiturus*) and grasswrens (*Amytornis*) are endemic to Australia. *Malurus* is found in a wide-range of habitats, but *Amytornis* in particular inhabits the arid Australian interior. The genus *Malurus* exhibits a biogeographic pattern in which the ranges of several sister species pairs do not overlap geographically, suggesting by some methods a history of allopatric speciation (Barraclough and Vogler 2000). The high level of plumage divergence among species and populations likely contributes to strong pre-mating isolating mechanisms, although in some clades, such as the largely allopatric chestnut-shouldered clade, plumage divergence is less extreme (Rowley and Russell 1997). If this history is accurate, then we can regard ILS rather than hybridization as the main source of heterogeneity in gene tree topologies in this group. Gene tree/species tree discordance due to both hybridization and ILS has been documented in many bird groups (Baker et al. 2003; McCracken and Sorenson 2005; Peters et al. 2007a, b), and paraphyly even of the rapidly sorting mtDNA genome has been detected in 44% of Australian birds (Joseph and Omland 2009; Joseph et al. 2009).

There have been several prior molecular phylogenetic studies on part or all of this family and they have used allozyme or DNA sequence data (Christidis and Schodde 1997; Christidis 1999; Christidis et al., 2010; Donnellan et al., 2009; Gardner et al. 2010). They mainly addressed phylogenetic relationships among species within and between the genera *Stipiturus*, *Amytornis* and *Malurus* as well as biogeographic patterns of strong association between species and habitats. These studies significantly contributed to the understanding of evolutionary relationships within the family, and to comparisons among tree-building methods. For example,

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Christidis et al. (2010) recently applied two mitochondrial and three nuclear loci to the phylogeny of grasswrens (*Amytornis*) and recovered substantial support within the group, somewhat higher with concatenation than with species tree methods. Still, many of the inferences made about relationships and biogeography remain controversial and more detailed molecular analyses are required. For example, it remains inconclusive as to whether the New Guinean and Australian fairy-wrens form separate monophyletic groups, and although the higher diversity of *Malurus* in northern Australia suggests a northern origin for this group, Christidis and Schodde (1997)'s results could not corroborate this result conclusively.

We applied species tree and traditional phylogenetic methods to further resolve evolution within the entire family of fairy-wrens and allied genera. Using eighteen different loci comprising four different types of genetic markers (anonymous, exonic, intronic, and indels), we were able to compare phylogenetic information and conflict within and among these groups of loci. We explore four different methods of estimating species trees: partitioned Bayesian analysis of concatenated sequences using Mr.Bayes 3.0 (Ronquist and Huelsenbeck 2003); Bayesian Estimation of Species Tree (BEST; Liu and Pearl 2007; Edwards et al. 2007); Species Tree estimation using Average Ranks of coalescences (STAR; Liu et al. 2009); and Bayesian Untangling of Concordance Knots (BUCKy; Ané et al. 2007). We discuss each of these methods in further detail below.

MATERIALS AND METHODS

Taxon Sampling and DNA Extraction

From four institutions we obtained tissue samples of 59 individuals representing 25 Australian Maluridae species and one outgroup (Appendix - Table 1). This coverage includes all currently recognized species of *Malurus*, *Stipiturus*, *Clytomias* and *Amytornis*, but does not

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

include the monotypic genus *Sipodotus*, for which no tissue samples were available. In contrast to previous phylogenetic studies of malurids, we included all members of the family in the ingroup. For the outgroup, we used one white-throated gerygone (*Gerygone albogularis*) from the Acanthizidae, which is closely allied to Maluridae within the Meliphagoidea (Driskell and Christidis 2004; Gardner et al. 2010). Although multiple outgroups are deemed superior in many phylogenetic studies, we used one outgroup due to the inability of several species tree methods to accommodate more than one outgroup sequence (Liu 2008; Liu et al. 2009). Genomic DNA was extracted from pectoral muscle samples using a standard genomic DNA extraction kit (Qiagen, Valencia, CA) and the manufacturer's protocol.

Development of Molecular Markers

A total of 18 genetic markers were used comprising four different types: five non-coding (as determined by BLAST; Lee and Edwards 2008) 'anonymous' nuclear loci (Mame-AL06; Mame-AL16; Mame-AL23; Mame-AL26; Mame-AL28), five introns (AB4, aldolase B intron 4; RI2, rhodopsin intron 2; TGFβ2, transforming growth factor-β2 intron 5; CDC132, coiled-coil domain containing 132; HMG-2, high mobility group protein B2), seven exons (FSHR, follicle stimulating hormone receptor; MEK1, MAP kinase-kinase; PTPN12, protein tyrosine phosphatase non-receptor type 12; TEX10, testis expressed gene 10; TNNT3, troponin T type 3; TRAF6, TNF receptor-associated factor 6; UBN1, ubinuclein 1) and a partial mitochondrial DNA gene (ND2, NADH dehydrogenase subunit 2; see details in Appendix - Table 2). Anonymous loci and introns are useful in phylogenetics of birds due to their high variability (Backström et al. 2008; Hackett et al. 2008), and exons have proven useful in higher level phylogenetic studies due to the low homoplasy and informative indels (e.g., Murphy et al. 2007). Primers for most of the genetic markers were obtained from previous studies of fairy-wrens or

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

other birds (Backström et al. 2008; Lee and Edwards 2008; Townsend et al. 2008), and the anonymous markers were aligned by blast to the zebra finch genome and chosen due to their lack of orthology with protein-coding regions (Lee and Edwards 2008). Anonymous loci and intron sequences from species also used in Lee and Edwards (2008) (n = 59 sequences) were used in this study. The primers for five exons (FSHR, MEK1, TEX10, TNNT3 and UBN1) were newly developed in this study. To design these primers, cDNA sequences of chicken and zebra finch for the five genes were downloaded from the GenBank and aligned in MacClade (Maddison and Maddison 2000). Using the UCSC genome browser (http://genome.ucsc.edu/), exon boundaries were located and primers were designed from the longest exon.

Amplification and Haplotype Estimation

All genetic markers were amplified in 20 µl reactions under the following conditions: denaturation at 94°C for 3 min followed by 35 cycles of 94°C for 30s, 60°C (55°C for all exons) for 30s, and 72°C for 1 min. This was followed by a 5 min extension at 72°C. Products were purified using a vacuum (Millipore, Billerica, MA), and subsequently a standard BigDye reaction (Applied biosystems, Foster City, CA) was performed. Products of these reactions were purified using the ethanol cleanup procedure and sequenced in both directions on an ABI 3100 (Applied Biosystems) Genetic Analyzer. DNA heterozygosity at two alleles of nuclear loci observed within single individuals continues to be ignored in many phylogenetic studies, presumably because such heterozygosity is thought to be absent or is considered not as useful as interspecific polymorphism in reconstructing phylogenetic relationships (see Groth and Barrowclough 1999 for an early acknowledgement of individual heterozygosity in an avian phylogenetic study). However, with the ability to incorporate into phylogenetic studies within-individual heterozygosity as well as between-individual polymorphism within species, analysis

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

of the two alleles that must comprise nuclear DNA sequence from each individual becomes essential. Ignoring heterozygosity tacitly assumes, among other things, that both alleles of an individual are most closely related to one another and therefore comprise a monophyletic group in gene trees. Therefore, we inspected individual chromatograms closely for putatively heterozygous sites, with evidence of two nucleotides, and scored these sites using the IUPAC code. Once such composite alleles from each individual were obtained, they were aligned in MacClade (Maddison and Maddison 2000), and haplotypes of genotypes that were heterozygous at multiple sites were inferred using the software PHASE2.1.1 (Stephens et al. 2001). Only those individuals for which PHASE was able to assign haplotypes with a probability greater than 0.70 were used in subsequent analyses. We did not attempt to detect recombination events within loci. To identify and resolve indels, we applied a technique involving manual inspection of chromatograms generated in both directions and subtraction of chromatogram peaks to identify indels and resolve haplotypes (Dolman and Moritz 2006; Lee and Edwards 2008). All sequences appearing for the first time in this paper have been deposited in the Genbank data base under accession numbers JN597307- JN598880. We here report the Genbank accession numbers used in Lee and Edwards (2008) since they were not reported in that study (FJ418984-FJ422117).

Throughout the study we used three subsets of the data for different analyses (Appendix - Table 1). When analyzing individual gene trees, we used all available sequences for all 26 species and 59 individuals (‘full data set’); these single-gene analyses included from 86 to 98 sequences. For multiple-allele, multiple individual species tree analyses, we reduced the full data set such that the individual-by-locus matrix was 100% complete. This resulted in a data set comprised of 42 ingroup individuals (84 alleles) across all 25 species (‘multi-allele data set’) and one outgroup allele. Finally, for some species tree analyses (see below) and for analysis of

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

congruence among loci, we used a one-sequence per species data set ('single-allele data set'), which was comprised of 26 sequences in total, and was also 100% filled (alleles pruned were randomly chosen). Analyzing the latter two data sets helped us to understand not only the effect of the number of alleles per species on the performance of a given species tree method but also to better understand the levels at which ILS may be occurring: With multiple alleles ILS between species can be distinguished form ILS among genera, whereas with single allele data set ILS can only occur among genera without assuming specific relationships among species within genera.

Estimation and Analysis of Gene Trees

Using the program MrModeltest 2.2 (Nylander 2004), sufficient DNA substitution models for each locus were estimated by the AIC test, and gene trees were estimated using MrBayes version 3.0 (Ronquist and Huelsenbeck 2003), with which we obtained Bayesian posterior probabilities from twenty million MCMC cycles with a sample frequency of 1,000 and a burn-in period of 2 million generations. Gene trees from MrBayes were used as input for the BUCKy analysis. For visualizing gene trees and for the STAR species tree analysis, we generated gene trees using Phyml v. 3.0 using the Generalized Time Reversible (GTR) model of nucleotide substitution and subtree pruning and regrafting searches (Guindon and Gascuel 2003). To assess the heterogeneity of gene trees and their underlying DNA sequences at species level, two programs were used: Concaterpillar 1.4 (Leigh et al. 2008) and Tree Set Visualization (TSV) module in Mesquite (Hillis et al. 2005). Concaterpillar uses phylogeny-based likelihood ratio tests to identify sets of statistically congruent and incongruent DNA sequence data sets, and is useful for identifying conflicting phylogenetic signal among loci (Leigh et al. 2008). TSV does not query DNA sequences but instead calculates the distance between every pair of gene trees (Robinson and Foulds 1981) and visualizes all gene trees as points in two-dimensional space

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

using multidimensional scaling (MDS). The TSV analysis was performed on both mrBayes and Phyml gene trees, as well as on gene trees generated using BEST, which uses a different prior on gene trees than mrBayes (see below). For these analysis we could not use the multi-allele data set because this would involve assumptions about which allele of each locus corresponded to alleles at other loci; we therefore used the single-allele data set for these tests.

Estimation of Species Trees

We used several methods for estimating species trees. First, we used BEST version 2.3 (Liu 2008), which has recently been used in several multilocus phylogenetic analyses of vertebrate groups (Belfiore et al. 2008; Leaché 2009). Several studies have shown that this approach is more accurate for estimating phylogenetic relationships than concatenation when internodes are very short and gene tree heterogeneity is high (Edwards et al. 2007; Kubatko and Degnan 2007; Belfiore et al. 2008). In BEST we obtained Bayesian posterior probabilities from sixy million MCMC cycles with a sample frequency of 1,000 and a burn-in period of 30 million generations, using a relatively flat prior for θ (α=3, β=0.03) since this has been shown to increase the rate of convergence in some data sets (Leaché 2009).

The BEST method requires intensive computation, rendering it less useful for large data sets; we were only able to achieve convergence using the single-allele data set. The STAR method (Liu et al. 2009) helps overcome this issue because of its fast use of gene tree summary statistics. In this method, rooted gene trees are first constructed for each locus using any kind of gene tree estimation method (e.g. Bayesian, maximum likelihood method, etc). Then, these gene trees are used to count the ranks between all pairs of species. (The rank of the coalescence at the root node is equal to the number of taxa in the tree, and then decreases by 1 as one moves from the root to the tips of the gene tree.) A distance matrix is made in which the entries are the

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

average ranks of coalescences in gene trees across loci, and this matrix can then be analyzed by a distance method, such as the neighbor-joining method. Loci and sites within loci can be bootstrapped so as to obtain confidence levels (Seo 2008). We implemented STAR using Phybase (Phylogenetic Analysis of multilocus sequences in R; Liu 2010), using 1000 multilocus bootstraps (Seo 2008) on both single- and multiple-allele data sets. We made a maximum likelihood gene tree for each bootstrapped dataset using Phyml as above. STAR trees from each bootstrapped set of gene trees were made using the neighbor-joining method in Phybase, and then a majority-rule consensus tree was made from these using the consensus function in Phybase. The single-allele matrix and species trees generated from this have been deposited in TreeBASE (accession number http://purl.org/phylo/treebase/phylows/study/TB2:S11849).

We also used the multilocus method BUCKy, which uses a Bayesian approach to estimate a 'dominant tree' comprised of those clades whose concordance factors, defined as the proportion of the genome for which a given clade is true, exceed those of any contradictory clades (Ané et al. 2007). Technically, this dominant tree is not necessarily the species tree. Bucky is agnostic as to the sources of gene tree heterogeneity, and instead of modeling ILS specifically it instead estimates a summary of the posterior distributions of the individual gene trees in a data set, and therefore will in some cases simply represent the most common branching pattern for a given clade across these gene tree distributions. Still, under many circumstances we expect this dominant tree to be similar to if not isomorphic with the species tree. We used the sub-program *mbsum* to summarize gene tree distributions generated for each locus in MrBayes and to perform Bayesian Concordance Analysis (BCA). All stored trees for each locus were used as input for *mbsum*. The output of *mbsum* was subsequently used for the subprogram *bucky*. Posterior probabilities for the dominant tree were obtained from five million MCMC cycles with

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

four different heating chains.

We also used the criterion of minimizing deep coalescences (MDC) to estimate a species tree (Maddison and Knowles 2006). We used the implementation of MDC in the computer package Phylonet (Than et al. 2008, 2009). Phylonet employs a dynamic programming method that results in fast and accurate determination of the MDC species tree from input files consisting of gene trees. We estimated unrooted gene trees using Phyml as above, then used these gene trees as input for Phylonet. We used the dyanamic programming option, estimating the MDC tree using unrooted gene trees. We incorporated uncertainty into our MDC tree by conducting 100 multilocus bootstrap replications of each 18-gene data set, then estimating gene trees using Phyml. The consensus of the 100 estimated MDC trees was then constructed using consensus function in Phybase.

Finally, we also took advantage of indels in our data set. All indels identified during manual investigation of chromatograms were checked again using DNAsp 5.0 (Librado and Rozas 2009), and subsequently recorded as binary characters in a separated file for each locus. We analyzed the concatenation of all indel characters using unweighted parsimony and 10 replicated heuristic searches using random addition of taxa in PAUP4.0a114 (Swofford 2003). We conducted bootstrapping on the indel data set alone and also reconstructed indels on the tree estimated on the concatenated data set of sequences.

Measuring the Level of Incomplete Lineage Sorting

To estimate the frequency of ILS across the Maluridae, we first assigned a discrete, unordered character state to each species in MacClade and counted the number of transitions of this character using unordered parsimony. These transitions are ‘interspecific coalescent events’ (Takahata 1989) and the ‘coalescent cost’ of gene trees (Flórez-Rodríguez et al. 2011) – the

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

number of deep coalescent events - should be minimized under complete reciprocal monophyly. Because this approach might miss instances of incongruence in which all species show reciprocal monophyly but incongruence exists deeper in the gene trees, we performed this on both single- and multiple allele data sets.

We next wanted to demonstrate that the 'messy' gene trees in our data are actually the result of ILS and not gene-tree estimation error (i.e., mutational error or stochasticity). We did this by asking if the signal in each gene could reject phylogenetic relationships in which there was no ILS. We used the Approximately Unbiased (AU) test (Shimodaira 2002) to test the hypothesis that the best gene tree for each locus (non-constrained gene tree) was a significantly better explanation of the data for that locus than was a gene tree in which all alleles were forced to be reciprocally monophyletic (constrained gene tree). We performed this test using the program consel (Shimodaira and Hasegawa 2001) with the single- and multi-allele data sets. In the multiple allele case we used a backbone of relationships among genera that matched the relationships in the species tree found in BEST; however, relationships of alleles within species were unresolved and treated as a soft polytomy so as not to assume any particular relationships among alleles within species. When using the single allele data set, we simply used the relationships of species found in the BEST tree. However by using the backbone of the BEST tree we assume that this tree is correct. To remedy this, we used a constraint tree in which the only constraint was monophyly of alleles within species – all other nodes were soft polytomies, yielding a star tree among all species and genera. We then found the maximum likelihood tree under this more relaxed constraint and then used the AU test to compare that tree with the best tree with no constraints. In addition, using the single allele data set allowed us to determine whether the signals for ILS were restricted to the trivial case of closely related sister species, or

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

whether they occurred deeper within the tree. The constrained gene trees were generated by finding the maximum likelihood tree under a particular constraint using GARLI version 0.951 (Zwickl 2006). We also tested phylogenetic trees in a concatenation framework using the AU test as well as the KH test of Kishino-Hasegawa (1989).

Results

Information Content of Molecular Markers

A total of ~1600 alleles were sequenced in this study. A total of 7,809 base pairs (bp) was sequenced for each individual, with an average length of 434 bp per locus (range 203 ~ 812; Table 1, Fig. 1). The ND2 region showed the highest variability and number of parsimony informative sites (Table 1). Non-coding nuclear loci (anonymous loci and introns) and exons exhibited levels of variation approximately half and a quarter, respectively, of the variation observed in ND2 (Table 1). A total of 102 indels were identified across all loci. Indels were found in all non-coding nuclear loci (mean frequency per locus = 9 ± 4; range 1 ~ 16; Table 1), and in three protein-encoding loci (1, 3 and 9 indels in loci TNNT3, MEK1 and UBN1, respectively). The one indel occurring in TNNT3 preserved the reading frame, whereas the other two loci experienced indels that disrupted the reading frame, making it likely that these coding regions are orphan exons or pseudogenes in the malurid genome.

Individual Gene Trees

All gene trees revealed a high level of heterogeneity in tree topologies and branch lengths, and no individual genes produced the same tree topology (Figs. 2 and 3). Although ND2 and some nuclear genes resulted in all or most species exhibiting reciprocal monophyly of alleles (e.g., Fig. 2A and B), we also found many instances of lack of reciprocal monophyly and putative ILS. Nonetheless, there were some general patterns in overall tree topology among the

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

loci. The monophyly of *Amytornis* grasswrens (*Amytornis*) was strongly supported by all gene trees and robustly positioned as sister to other genera in the family. Although the *Stipiturus* emu-wrens were also monophyletic in all gene trees, the placement of individual species and the relationships of the clade varied across loci (Fig. 3). *Malurus* was the least stable of the genera, and five gene trees failed to recover monophyly of the 'core malurids' (all *Malurus* except *Clytomyias* and *M. grayi*).

Counting interspecific coalescent events by parsimony (Methods), we found some genes that exhibited up to 11 extra steps (e.g., TEX10, Fig. 3C; Table 1). Close inspection of gene trees showed frequent lack of reciprocal monophyly within *Malurus, Amytornis and Stipiturus* (Figs. 3 and 4) as well as variation in gene trees among genera (Fig. 2D-F). We visualized gene tree variation using TSV (Fig. 4A-C) and found that there was a wide range of gene trees in multidimensional space. The RF distances among gene trees were substantially smaller for the BEST analysis than for either Phyml or mrBayes (Fig. 4D; see below).

Concaterpillar revealed that there were three statistically distinct groups of loci at the DNA sequence level (Fig. 4): [AL06, AL28, AB4, HMG-2, TGFβ2], [AL16, AL23, AL26, ND2], and [FSHR, MEK1, PTPN12, TEX10, TNNT3, TRAF6, UBN1, CDC132, RI2]. Surprisingly, three of the locus types (anonymous, intron and exon) showed detectable within-group phylogenetic compatibility: for example, all of the exons fell into one group and three of five introns and anonymous loci fell into their own groups (Fig. 4).

Species Trees

Single allele dataset.--The hierarchical Bayesian method (BEST) yielded a well-resolved tree with many nodes exhibiting high posterior probability values when using the single-allele data set (17 out of 23 ingroup nodes had posterior probabilities ≥ 0.95; Fig. 5). The consensus

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

species tree showed that *Amytornis* was sister to all other genera, followed by divergence of *Malurus* and *Stipiturus*. The other three species tree methods generated broadly similar results, as did concatenation (Fig. 6). All four methods recovered monophyly of each genus and assigned *Amytornis* as sister to all others. We found two major discordances in tree topology among methods, however. First, across the four methods, relationships varied among the species of the chestnut-shouldered group of four fairy-wren species, (*Malurus amabilis*, *M. lamberti*, *M. elegans*, and *M. pulcherrimus*), and also among species in a clade comprising *Amytornis ballarae*, *A. purnelli*, *A. housei* and *A. goyderi*. Furthermore, the placement of *M. coronatus* varied among trees: it was either sister to *M. alboscapulatus* + *M. leucopterus* + *M. melanocephalus* (BEST) or to a broader clade comprising those three and *M. cyaneus* + *M. splendens* (STAR, BUCKy). Concatenation gave generally higher confidence for each node and BUCKy had relatively low concordance factors for each node overall. MDC produced a species tree that was somewhat divergent from the others in placing *M. coronatus* as sister to *M. cyanocephalus* (albeit with low confidence) and *M. lamberti* as sister to the remaining chestnut-shouldered group.

Using TSV, we plotted the species trees in multidimensional space along with the 18 gene trees (Fig. 4). The five species trees in Figs. 5 and 6 clustered closer to one another than to any individual gene tree (Fig. 4A, B), except for in the BEST analysis, where gene and species trees were more intermingled (Fig. 4C). The Robinson-Foulds distances bear this out, with average RF distances between gene trees and all species trees being larger for mrBayes and Phyml (21.7 ± 5.0 [1 s.d.] and 22.5 ± 5.8, respectively) than for BEST (12.1 ± 4.1). With both mrBayes and Phyml, the five species trees fell into a distinct cluster somewhat separate from the constituent gene trees (Fig. 4A,B; see below). By contrast, in the BEST analysis, the gene trees and species

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

trees were more intermingled (Fig. 4C). The gene trees that fell closest to estimated species trees in multidimensional space differed for Phyml and mrBayes. Whereas with Phyml two anonymous loci, AL16 and AL23, were closest to the main cluster of species trees, with mrBayes, the ND2 tree fell closest to the species tree. In Figs. 4A and B, exon gene trees clustered farther away from the multilocus phylogenies, whereas gene trees of anonymous loci had the highest variation in clustering position among locus types.

Multiple-allele dataset.--Here we were able to obtain species trees only from two methods: BUCKy and STAR. We did not analyze the multiple allele data set with concatenation because it was not clear which alleles to concatenate for different loci. Six independent runs of the multiple allele dataset using four different heating schemes in BEST failed to converge after 300 million MCMC cycles, each run of which took several weeks, even when clearly monophyletic groups were analyzed separately (log likelihood values after 250 million cycles: -40690.452, -44577.431, -48430.936, -39640.443, -39317.347, -40435.580). STAR and BUCKy showed similar phylogenies as above, but support and concordance values were generally lower than in the results for the one-allele dataset (Figure 6B and C).

Indel phylogeny.--We encountered 102 indels across the multi-allele data set (84 alleles; Table 1). Of the 102 indels, 80 were parsimony informative. Unordered parsimony analysis on the entire indel data set was based on a maximum of 5000 equally parsimonious trees saved and resulted in a set of trees of 129 steps with consistency indexes of 0.79. These trees (not shown) revealed relationships broadly similar to those in Figs. 5 and 6 albeit with several species paraphyletic or unresolved: *A. ballarae, housei, M. amabilis, elegans, lamberti, cyaneus, melanocephalus, leucopterus* and *C. insignis*. In these trees *A. dorotheae* and *striatus* were sister taxa as in the concatenated sequence tree (Fig. 6A) and *M. cyanocephalus* fell outside the clade

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

consisting of emu wrens (*Stipiturus*) and the remaining Malurinae. Across all 102 indels, 79 are inferred to have changed once, 19 changed twice and four changed three times. There were a total of 32 unambiguously reconstructed deletions and 13 insertions on this tree. We reduced the number of OTUs in the indel data set so as to compare with the other trees using the single-allele data set (25 alleles). This reduced data set had 24 invariant and 48 parsimony informative characters as a result of the taxon deletion, and resulted in 120 equally parsimonious trees of length 95 and consistency indexes of 0.82 (Fig. 7). Although there were several unresolved nodes, the indel tree recovered several major splits among genera that were also in the other trees. In particular, the relationships recovered within *Stipiturus* were the same as in other methods. Across all variable indels, 63 changed once on the tree, 13 changed twice, with two changed three times. Plotting of unambiguously reconstructed indels on this tree revealed a number of phylogenetically informative events throughout *Amytornis* and to a lesser extent within *Malurus*, particularly on the branch leading to the chestnut-shouldered group (Fig. 7). Overall on this reduced data set there were 41 deletions and 10 insertions that could be unambiguously reconstructed.

Measuring Incomplete Lineage Sorting via Phylogenetic Signal

We also tested for the presence of ILS by asking if the sequence data for a given locus could reject species trees using the AU test. We first applied the AU test to the full data set with the maximum number of alleles constrained to be monophyletic for each locus (full data set) and using the backbone provided by the BEST species tree. In this case, every locus except for the intron AB4 could reject the constraint tree in which all alleles were monophyletic within species, implying a substantial amount of ILS across the tree. At 203 bp, AB4 is the shortest locus in our data set and therefore may not have the power to adequately test for ILS. To test this hypothesis,

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

we pseudoduplicated and triplicated the locus and repeated the AU tests. With a double concatenation, the *p*-value of the AU decreased to 0.058, and when triplicated it decreased further to 0.033, suggesting that the little variation in the locus may tend towards rejecting reciprocal monophyly.

AU tests applied to the single allele data set revealed that 12 out of 18 loci could reject a constrained tree, suggesting the presence of ILS above the species level. Six of the 18 (AL06, FSHR, TRAF6, UBN1, HMG-2, and ND2) could not reject the constraint tree, and we accept the null hypothesis of consistency with the species tree. Surprisingly, these six loci were not depauperate in variation – they do not rank among the lowest in terms of the number of phylogenetically informative sites (Table 1; Fig. 1). Thus the information content in these loci appears to trend towards congruence with the species tree. Because the BEST tree may not be the correct tree, we repeated the AU tests with the full data set but this time using a constraint tree whose only constraint was monophyly of alleles within species; the relationships among species in the gene trees was unconstrained and allowed to vary by locus. We found that all genes except for MEK1 and ND2 could reject even this relaxed constraint, again suggesting substantial ILS.

Finally, we used the AU test to see test the phylogenetic consistency of the entire concatenation of the single-allele data set with the five species tree methods (mrBayes, BEST, STAR, BUCKy and MDC; Table 2). In this setting, the tree produced by Phyml had the highest likelihood. Under this framework, it is no surprise that trees generated by concatenation should be ranked higher than trees built by species tree methods; AU tests in a species tree framework have not yet been developed (but see Carstens and Knowles 2007; Knowles and Carstens (2007). We found that the concatenated data set could not reject any concatenated tree or species tree

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

except for those produced by MDC. This divergence of the MDC tree is reflected also in its divergence placement in multidimensional space in the TSV analysis (Fig. 4A, B).

Discussion

Molecular Markers in Avian Phylogenetics

We have used four different types of genetic markers - anonymous loci, introns, exons and mtDNA – to resolve the evolutionary history of one of the most iconic bird families in Australia and New Guinea, the fairy-wrens and allies. Using tests of phylogenetic signal for each of our 18 loci, we have demonstrated that incomplete lineage sorting (ILS) is a genuine feature of our data set, as opposed to noise generated from mutational effects of lack of phylogenetic signal (Huang et al. 2010). ILS occurred both between closely related species as well as deeper in gene trees among genera, suggesting that ILS is likely a general feature of the genetic history of avian species and genera. Overall the level of gene tree heterogeneity in our data set was high, suggesting that the confidence of nodes the tree made by concatenation may be spuriously high. The four species tree methods we used (BEST, STAR, BUCKy, MDC) generated estimates of phylogeny that were generally congruent and similar to that produced by concatenation, although MDC produced species trees that were more divergent from those of other methods in multidimensional space. Overall main phylogenetic results support the recent findings of Christidis et al. (2010) for *Amytornis;* suggest rapid evolution and high levels of ILS in the chestnut-shouldered group of *Malurus*; and show conclusively that the Australian and New Guinea species of *Malurus* do not form separate monophyletic groups.

The diversity in the topologies of the gene trees produced by mrBayes, Phyml and BEST varied among tree building methods. The RF distance among gene trees was significantly smaller for gene trees produced in BEST as compared to gene trees produced in mrBayes or

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Phyml (Fig. 4). This result likely stems from the different prior used by BEST; whereas BEST assumes that gene trees are correlated among themselves and with the species tree, standard phylogenetic methods make no such assumption or assume a flat prior on gene trees. Biologically, it is more plausible that gene trees should be correlated with one another due to the correlation imposed by the species tree (Liu and Pearl 2007; Edwards et al. 2007). The multispecies coalescent ensures that some variation in gene tree topologies will occur despite the increased similarity among gene trees (Degnan and Rosenberg 2009). In addition, the average RF distance between the 18 gene trees and each of the species trees was substantially smaller in the BEST analysis as compared to gene trees made with mrBayes or Phyml (Fig. 4). This result also likely stems from the more concentrated posterior distribution of gene trees and the use of the joint prior in BEST.

The anonymous loci we used in our study possessed substantial variation but, as found in other studies (e.g., Thomson et al. 2008), their utility declined with phylogenetic distance from the species from which primers were designed. We originally tested primers from the same 29 anonymous loci markers used in a previous phylogeographic study on *Malurus melanocephalus*, the Red-backed Wren (Lee and Edwards 2008), yet we found that only 9 loci (~31%) amplified a single band and produced useable sequences for all malurids. Anonymous loci are useful for species-level studies (Lee and Edwards 2008; Balakrishnan and Edwards 2009; Brito and Edwards 2009), but at higher taxonomic levels, finding appropriate anonymous markers becomes difficult. By contrast, because intron primers are usually developed from conserved flanking exons, they can be applied more easily to a wider range of organisms (Backström et al. 2008), and in our study they provided as much phylogenetic information as anonymous loci. Furthermore, Chojnowski et al. (2008) found that introns provided more resolution for basal

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

branches in the tree for birds than did exons for clathrin heavy chain genes. Overall, therefore, intron markers may be the most efficient approach to studying phylogenetic relationships at higher levels in birds, although more studies are needed.

Comparing Methods of Species Tree Estimation

As the importance of applying a multi-locus approach to phylogenetic studies becomes increasingly recognized, and as appropriate methods become more accessible, it is important to evaluate relative performance of different species tree estimation methods (Belfiore et al. 2008; Brumfield et al. 2008; Linnen and Farrell 2008). Overall, the five methods generated similar results in terms of tree topology, but we were unable to compare branch lengths because three methods (BUCKy, STAR and MDC) do not provide branch length outputs. There are two clades, one in *Malurus* and the other in *Amytornis*, at which we found discordance across the four species tree estimation methods we have compared (Figs. 5 and 6). Whereas BEST assigned low probabilities of confidence on these clades, concatenation placed >90% confidence on those two clades. Several studies have noted the discrepancy between posterior probabilities of concatenation and species tree methods (Belfiore et al 2008; Thomson et al. 2008; Brumfield et al. 2008; Leaché 2009), and our study confirms these trends. The low resolution at some clades may represent the reality that these nodes require larger amounts of data to effectively resolve. On the other hand, the low confidence at these nodes may seem surprising given the size of our data set. Our emphasis was on increasing the number of loci rather than individuals per species, given the clear improvements that larger numbers of loci confer on species tree estimation (Maddison and Knowles 2006; Edwards et al. 2007; Liu et al. 2009; McCormack et al. 2008; Leaché and Rannala 2011). But the small number of individuals per species used in this study might be driving low resolution at some nodes. Sampling larger numbers of individuals is

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

advantageous for some species tree methods, especially when lineage lengths are short in coalescent units (Maddison and Knowles 2006; McCormack et al. 2008). It will be useful to confirm the influence of individual number on species tree estimation using additional methodologies. An additional factor lowering our confidence in the estimated species tree might be recombination within species, which we ignored in this study. Recombination could not only affect sequences within species but could in principle occur among allelic lineages in common ancestral lineages in the tree (Jennings and Edwards 2005). Recombination is known to decrease the efficiency of BEST and likely compromises other species tree methods as well (Castillo-Ramírez et al. 2010).

BUCKy also allows gene trees to have different genealogies, and rather than employing a coalescent model, it is agnostic as to the source of gene tree variation. The summary tree produced is an attempt at minimizing and summarizing gene tree discordance and may not necessarily represent the species tree. It would be useful to test BUCKy on gene trees generated from the anomaly zone, a zone of species tree space in which the most common gene tree is discordant with the true species tree (Degnan and Rosenberg 2006). Whereas BEST has been shown to be consistent in the anomaly zone (Liu and Edwards 2009), it may be that BUCKy yields an incorrect tree that favors the most common gene tree. This of course is not a criticism of BUCKy given that its purpose is to summarize genomic variation rather than explicitly to estimate species trees.

Like the other species tree methods, STAR relies on coalescent theory, but it does so through summary statistics, in this case the average ranks of pairs of species (Liu et al. 2009). Unlike the other three methods, STAR ultimately utilizes a distance approach on gene trees rather than a Bayesian approach, although the gene trees that it uses as input can be estimated by any method,

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

including Bayesian methods. Like BUCKy, STAR does not estimate branch lengths (Liu et al. 2009). Nonetheless, it generates a phylogeny whose topology is closer to that estimated by BEST than to a tree inferred by MDC (see supplementary material available from http://www.sysbio.oxfordjournals .org/]). Although STAR assumes that incongruence between gene and species trees is exclusively due to deep coalescence, and is sensitive to some types of model violations such as genome-wide introgression, it is nonetheless very robust to other violations of the coalescent model involving individual genes, such as horizontal gene transfer (Liu et al. 2009). The MDC method produced a tree with two somewhat anomalous relationships (see results), causing it to deviate from the other methods in tree space using TSV. This TSV method queries only the tree topology and not the sequence data. Even so, AU tests across the entire 18-gene data set showed that the concatenated sequences rejected the topology produced by MDC, yet could not distinguish between the result from mrBayes and the four other species tree methods. It is known that MDC can be statistically inconsistent in some situations, such as in the anomaly zone; whether or not this zone is driving the divergent MDC results here is not known.

Finally, we used indels to infer phylogenetic relationships. Indels are not universally observed to be free of homoplasy (e.g., Belinky et al. 2009; Churakov et al. 2010), and, like other non-SNP genomic variation such as retroposons (Shedlock et al. 2004) are also subject to ILS. Nonetheless we found that the 102 indels in our data set were very informative, displayed low consistency indexes and were able to recover monophyly for each genus on their own. There was a clear bias towards deletions: depending the data set analyzed, the bias towards deletions among unambiguously reconstructed indels was ~2.5 (multi-allele data set) or 4.2. Johnson (2004) surveyed phylogenetic variation in intron 7 of the β-fibrinogen gene of pigeons and doves

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

and estimated a deletion bias of ~6. There are too few comparisons to tell if the bias we have detected in our data set is significantly different from that found in other birds. What is noteworthy however is the consistent estimate of a deletion bias in the few avian studies thav have examined this question. The instances of homoplasy in the indel data set could represent ILS or, alternatively, true homoplasy, two hypotheses that we cannot easily distinguish at this time. Despite the fact that detailed relationships within each genus are less well resolved than for sequence data or differ slightly from our consensus species tree, our study confirms earlier work that indels, particularly in noncoding loci, are an important source of phylogenetic information in birds (e.g., Ericson et al. 2000) and other taxa (Bardeleben et al. 2005; Matthee et al. 2007; Lake et al. 2008).

Substantial Level of Incomplete Lineage Sorting

Our analysis suggests that ILS is common in our data set, potentially affecting every locus. If ILS is the sole cause of incongruence between gene and species trees, and if our gene and species trees are reconstructed accurately, our analysis suggests that ILS occurs not only between closely related species but deeper in the phylogeny as well. Although the AU test does not provide information on where in the gene tree ILS occurs, our use of the AU test on our single-allele data set means that any incongruence between gene and species trees must occur between non-sister species, because only one allele per species was sampled. In fact, we did find several instances of incongruence occurring even between genera in our study. For example, in several genes the alleles from emu-wrens (*Stipiturus*) cluster most closely with those from basal malurids (*C. insignis* and *M. grayi*; e.g., AL28) or more closely to the core *Malurus* clade than to the base of *Malurus* (AL16, AL26). We have mentioned that some of this discordance among the deeper nodes in the gene trees (e.g., among *Amytornis* sp. for the ND2 gene and among major

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

1
2
3 groups for other genes) is not statistically significant and is no doubt due to poor resolution or incorrect phylogenetic reconstruction. Some discordance could in principle be caused by hybridization. If hybridization is present it would constitute a violation of several species tree methods, including BEST and STAR. Although extensive gene flow can indeed be problematic for species tree methods (Eckert and Carstens 2008), several studies have reported reasonable results if the taxa exchanging genes are closely related (e.g., Brumfield et al. 2008; Brumfield and Carling 2010). However, the distributions of nearly all clades in this study are strongly allopatric; indeed, *Malurus* has been a model for the inference of allopatric speciation throughout the entire genus based on phylogenetic analysis (Barraclough and Vogler 2000), and present distributions in *Amytornis* are also strongly allopatric, although some have argued for sympatry or parapatry between some *Amytornis* in the recent past (Black 2004). Thus any hybridization that would have occurred in *Amytornis* most likely would have been earlier rather than later in the history of the group. In general, given the strong pre-mating isolating mechanisms (at least in *Malurus*) and the strongly allopatric distributions in the clade, we suspect that hybridization is unlikely to have given rise to substantial ILS patterns in this study.

Additionally, incongruence at deep nodes due to ILS in birds should not come as a surprise to avian systematists, given the high incidence of shared alleles observed among species and even genera of birds in many allozyme studies (reviewed in Avise and Aquadro 1982; Balakrishnan et al. 2010) and increasingly in DNA studies of birds and other taxa (Jennings and Edwards 2005; Pollard et al. 2006).

Systematics of the Maluridae

The systematic implications of our results for *Amytornis* and *Stipiturus* are consistent with Christidis et al. (2010), Gardner et al. (2010) and Donnellan et al. (2009) especially the

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

placement of *A. housei, A. goyderi, A., ballarae* and *A. purnelli* as closest relatives. Using allozyme analysis, Christidis (1999) showed that *A. barbatus* is the earliest lineage to have diverged in *Amytornis*, a result confirmed by a recent DNA sequencing study of the genus (Christidis et al. 2010). Our sequence-based analysis confirms this general pattern (*A. barbatus* was sister to all other *Amytornis*), but suggests additionally that the *striatus* and *textilis* complexes as earlier construed (e.g., Schodde 1982) are not monophyletic.

The phylogenetic position of monotypic *Clytomyias* of New Guinea has been especially uncertain, however. It has been aligned with *Stipiturus* or left unresolved (Christidis and Schodde 1997) and limited taxon sampling has limited the power of other studies to resolve its position (Gardner et al. 2010). We find *Clytomyias* to be the sister of *M. grayi*, which is another New Guinean endemic. Together they form a clade that is sister to all other *Malurus*. This renders *Malurus* paraphyletic. Given that *M. grayi* is itself phenotypically most unlike other *Malurus* species (Schodde 1982), several options arise for eliminating paraphyly of *Malurus*: (1) expand *Malurus* to accommodate *M. grayi* and *C. insiginis*, (2) combine *grayi* and *insignis* in *Chenoramphus* Oustalet, 1878, which has priority over *Clytomyias* Sharpe, 1879 and in which *grayi* has been placed, (3) retain monotypic *Clytomyias* and reinstate *Chenoramphus* for *M. grayi*. We advocate the third option given the proviso that one malurid species *Sipodotus wallacii* still remains absent from DNA sequence-based molecular data sets.

Finally, *Malurus* comprises three major groups: the chestnut-shouldered group, the bi-colored group, and the blue group (Rowley and Russell 1997). This morphological categorization is generally congruent with our molecular data: the first two groups are monophyletic but the blue group is not. The apparent discordance is in the relationships within the bi-color group. Previously, *M. melanocephalus* had been grouped with either *M. leucopterus* or *M.*

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

alboscapulatus. Our results suggest that *M. alboscapulatus* is sister to *M. melanocephalus*, not *M. leucopterus*. Notably, this provides a more parsimonious explanation for reverse sexual dimorphism that we observe in this bi-colored group. *M. alboscapulatus* and *M. melanocephalus* exhibit reverse sexual dimorphism where the tail of males significantly shortens during the breeding season such that it becomes even shorter than the tail of females It has been hypothesized that this odd sexual polymorphism arose twice independently, given that *M. leucopterus* was thought to be closer to *M. melanocephalus* (Swaddle et al. 1999). However, our results show that this unexpected evolutionary pattern can be explained with a single origin in this group (note that it also occurs in *Clytomyias* Swaddle et al. 1999).

There is the evidence of rapid evolution in the chestnut-shouldered group and the topology we show among these species is subject to further testing and resolution. As suggested by the short internal node, this group may have gone through rapid speciation events. This hypothesis is further supported by the four species comprising the chestnut-shouldered group being mostly allopatric (*M. lamberti* and *M. pulcherrimus* have limited overlap) but phenotypically similar (Ford 1966; Schodde 1982). The nuclear gene trees show that this group and related *Malurus* show substantial ILS (Figs. 2 and 3), a pattern often found in rapidly speciating bird clades (Baker et al. 2003; McCracken and Sorenson 2005; Joseph et al. 2009). Further detailed study of gene trees and phylogenetic relationships among these species is warranted.

Biogeographic Inferences

A central Australian arid zone origin for *Amytornis* was advocated by (Keast 1961; Ford 1974, 1987) whereas Schodde (1982) origins in the northern monsoon region. The latter was based on the greater genetic differentiation of northern tropical species compared to those in the center of the continent. However, the distinction between members of what he considered to be

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

the *textilis* group, for example, as being central and the *striatus* group (*A. woodwardi*, *A. striatus* and in his analysis *A. merrotsyi*) being northern is not clear-cut. Our trees generally support a central Australian origin of *Amytornis*.

With *Malurus* circumscribed as recommended above, we have affirmed that Australian and New Guinean species of *Malurus* do not form separate, monophyletic groups. Given the basal phylogenetic positions of New Guinea endemics *Clytomyias*, *Chenoramphus*, and *M. cyanocephalus*, one might argue an origin of *Malurus* there. However, based on the absence of chestnut-shouldered fairy-wrens in wetter temperate, south-east Australia and the variety of representatives in tropical, arid and subtropical northern Australia, a northern origin has been hypothesized for this group (Schodde 1982; Rowley and Russell 1997). Clearly a more quantitative biogeographic analysis is warranted. Hopefully the phylogenetic hypotheses presented here will facilitate such an endeavor.

Note added in copy edit.--Driskell et al. (2011) have recently published in a short communication a phylogeny of the Malurinae (*Malurus, Sipodotus, Chenorhamphus* [=Malurus] *grayi* and *Clytomais* using four mitochondrial genes and three nuclear markers. We refer the reader to that paper to discern similarities and differences from our study.

FUNDING

Laboratory work was supported by National Science Foundation grant DEB – 0108249 (0500862) to SVE and a National Science Foundation Doctoral Dissertation Improvement Grant 0909897 to JYL and SVE.

ACKNOWLEDGEMENTS

We thank Jessica Leigh for help with concaterpillar, Liang Liu for help with Phybase and STAR, Yun Yu and Luay Nakhleh for help with Phylonet, Hidetoshi Shimodira for help with consel,

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Terri Kaltz for help installing concaterpillar on the Harvard FAS Research Computing Cluster, and Frank Rheindt for helpful discussion and comments on an earlier draft of the manuscript. We thank David Swofford for pointing out (after our paper was accepted) that technically we misapplied the AU test because we compared trees optimized by the data rather than trees chosen *a priori*. Two anonymous reviewers, Robb Brumfield and Ron DeBry improved the manuscript. We thank Robert Palmer and Anne Hastings (CSIRO Ecosystems Sciences) for sending samples from the collection of the ANWC and assistance with figures, respectively. Peter Marsack drew and provided the images of birds in Fig. 5 (originally published in Rowley and Russell 1997). Tyrone Li helped submit the sequences to Genbank. All specimens studied were collected under appropriate ethics and scientific collecting permits granted by the following agencies or their precursors CSIRO Ecosystems Sciences, Queensland Environmental Protection Agency, Queensland Parks and Wildlife Service; Biodiversity Management Unit, National Parks and Wildlife Service, New South Wales; National Parks and Wildlife Service, Department of Environment and Heritage, South Australia; Department of Environment and Conservation, Western Australia, Department of Natural Resources, the Environment, The Arts and Sport, Northern Territory, Conservator of Fauna and Department of Environment and Conservation, Papua New Guinea.

References

Ané, C., B. Larget, D. A. Baum, S. D. Smith, and A. Rokas. 2007. Bayesian estimation of concordance among gene trees. Mol. Biol. Evol. 24:412-426.

Avise, J.C., and C.F. Aquadro. 1982. A comparative summary of genetic distances in the vertebrates – patterns and correlations. Evol. Biol. 15: 151-185.

Avise, J. C. 2000. *Phylogeography*. Harvard University Press, Cambridge, MA.

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Backström, N., S. Fagerberg, and H. Ellegren. 2008. Genomics of natural bird populations: a gene-based set of reference markers evenly spread across the avian genome. Mol. Ecol. 17:964-980.

Baker, J. M., E. Lopez-Medrano, A. G. Navarro-Siguenza, O. R. Rojas-Soto, and K. E. Omland. 2003. Recent speciation in the orchard oriole group: Divergence of *Icterus spurius spurius* and *Icterus spurius fuertesi*. Auk 120:848-859.

Balakrishnan, C. N., and S. V. Edwards. 2009. Nucleotide variation, linkage disequilibrium and founder-facilitated speciation in wild populations of the zebra finch (*Taeniopygia* guttata). Genetics 181:165-180.

Balakrishnan, C.N., Lee J. Y. and Edwards, S. V. 2010. Phylogeography and phylogenetics in the nuclear age. In press in: *Searching for the Causes of Evolution: From Field Observations to Mechanisms*. (Eds., Peter and Rosemary Grant), Princeton University Press, Princeton, NJ.

Bardeleben, C., R. L. Moore, and R. K. Wayne. 2005. A molecular phylogeny of the Canidae based on six nuclear loci. Mol. Phylogenet. Evol. 37:815-831.

Barraclough, T. G., and A. P. Vogler. 2000. Detecting the geographical pattern of speciation from species-level phylogenies. Am. Nat. 155:419-434.

Belfiore, N. M., L. Liu, and C. Moritz. 2008. Multilocus phylogenetic of a rapid radiation in the genus *Thomomys* (Rodentia: Geomyidae). Syst. Biol. 57:294-310.

Belinky, F., O. Cohen, and D. Huchon. 2010. Large-Scale Parsimony Analysis of Metazoan Indels in Protein-Coding Genes. Mol. Biol. Evol. 27:441-451.

Black, A.B. 2004. The 'Immarna' grasswrens of R.C. Chandler: locality, habitat, identity and taxonomic implications. S. Aust. Ornith. 34: 199-211.

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

Brito, P. H., and S. V. Edwards. 2009. Multilocus phylogeography and phylogenetics using sequence-based markers. Genetica 135:439-455.

Brooker, M. G., I. Rowley, M. Adams, and P. R. Baverstock. 1990. Promiscuity: an inbreeding avoidance mechanism in a socially monogamous species. Behav. Ecol. and Sociobiol. 26:191-200.

Brumfield, R. T., and S. V. Edwards. 2007. Evolution into and out of the Andes: a Bayesian analysis of historical diversification in *Thamnophilus* antshrikes. Evolution 61:346-367.

Brumfield, R. T., L. Liu, D. E. Lum, and S. V. Edwards. 2008. Comparison of species tree methods for reconstructing the phylogeny of bearded manakins (Aves: Pipridae: *Manacus*) from multilocus sequence data. Syst. Biol. 57:719-731.

Brumfield, R.T. and M.D. Carling. 2010. The influence of hybrid zones on species tree inference in manakins. Pp. 115-128 in Estimating Species Trees: Practical and Theoretical Aspects (L.L. Knowles and L.S. Kubatko, eds.) Wiley-Blackwell, Hoboken, NJ.

Carstens, B. C., and L. L. Knowles. 2007. Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from melanoplus grasshoppers. Syst Biol 56:400-411.

Castillo-Ramírez, S., L. Liu, D. Pearl, and S. V. Edwards. 2010. Bayesian estimation of species trees: a practical guide to optimal sampling and analysis. Pages 15-33 in Estimating Species Trees: Practical and Theoretical Aspects (L. L. Knowles, and L. S. Kubatko, eds.). Wiley-Blackwell, New Jersey.

Chojnowski, J. L., R. T. Kimball, and E. L. Braun. 2008. Introns outperform exons in analyses of basal avain phylogeny using clathrin heavy chain genes. Gene 410:89-96.

Christidis, L. 1999. Evolution and biogeography of the Australian grasswrens, *Amytornis* (Aves:

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

Maluridae): biochemical perspectives. Aust. J. Zool. 47:113-124.

Christidis, L., and R. Schodde. 1997. Relationships within the Australo-Papuan fairy-wrens (Aves: Malurinae): an evaluation of the utility of allozyme data. Aust. J. Zool. 45:113-129.

Christidis, L. and Norman, J. A. 2010. Evolution of the Australasian songbird fauna. Emu 110: 21-31.

Christidis, L., Rheindt, F. E., Boles, W. E., and Norman, J. A. 2010. Plumage patterns are good indicators of taxonomic diversity, but not of phylogenetic affinities, in Australian grasswrens *Amytornis* (Aves: Maluridae). doi:10.1016/j.ympev.2010.08.029

Churakov, G., M. K. Sadasivuni, K. R. Rosenbloom, D. Huchon, J. Brosius, and J. Schmitz. 2010. Rodent evolution: back to the root. Mol. Biol. Evol. 27:1315-1326.

Degnan, J. H., and N. A. Rosenberg. 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2:762-768.

Degnan, J. H. and Rosenberg, N. A. (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends in Ecology & Evolution 24:332-340.

Donnellan, S.J., Armstrong J., Pickett, M., Milne, T., Baulderstone, J., Hollfelder, T., and Bertozzi, T. 2009. Systematic and conservation implications of mitochondrial DNA diversity in emu-wrens, *Stipiturus* (Aves: Maluridae). Emu 109: 143–152.

Driskell, A. C., and L. Christidis. 2004. Phylogeny and evolution of the Australo-Papuan honeyeaters (Passeriformes, Meliphagidae). Mol. Phylogenet. Evol. 31:943-960.

Driskell, A. C., J. A. Norman, S. Pruett-Jones, E. Mangall, S. Sonsthagen, and L. Christidis. 2011. A multigene phylogeny examining evolutionary and ecological relationships in the Australo-papuan wrens of the subfamily Malurinae (Aves). Molecular Phylogenetics and

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

Evolution 60:480-485.

Eckert, A.J., Carstens, B.C., 2008. Does gene flow destroy phylogenetic signal? The performance of three methods for estimating species phylogenies in the presence of gene flow. Mol. Phylogenet. Evol. 49, 832–842.

Edwards, S.V. 2009. Is a new and general theory of molecular systematics emerging? Evolution 63: 1-19.

Edwards, S. V., L. Liu, and D. K. Pearl. 2007. High-resolution species trees without concatenation. Proc. Natl. Acad. Sci. (USA) 104:5936-5941.

Ericson, P. G. P., U. S. Johansson, and T. J. Parsons. 2000. Major divisions in oscines revealed by insertions in the nuclear gene c-myc: a novel gene in avian phylogenetics. Auk 117:1069-1078.

Felsenstein, J. 2004. Inferring Phylogeny. Sinauer Associates, Sunderland, Massachusetts.

Florez-Rodriguez, A., M. D. Carling, and C. D. Cadena. 2011. Reconstructing the phylogeny of "*Buarremon*" brush-finches and near relatives (Aves, Emberizidae) from individual gene trees. Molecular Phylogenetics and Evolution 58:297-303.

Ford, J. 1966. Taxonomy and variation of the chestnut-shouldered wrens of Western Australia. Emu 66: 47-57.

Ford, J. 1974. Speciation in Australian birds adapted to arid habitats. Emu 74:161-168.

Ford, J. 1987. Minor isolates and minor geographical barriers in avian speciation in continental Australia. Emu 87:90-102.

Fujita, M. K., J. A. McGuire, S. C. Donnellan, and C. Moritz. 2010. Diversification and persistence at the arid-monsoonal interface: Australia-wide biogeography of the Bynoe's gecko (*Heteronotia binoei*; Gekkonidae). Evolution 64: 2293–2314.

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

Funk, D. J., and K. E. Omland. 2003. Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. Ann. Rev. Ecol. Syst. 34: 397-423.

Gardner, J., Trueman, J., Ebert, D., Joseph, L. and Magrath, R.D. 2010. Phylogeny and evolution of the Meliphagoidea, the largest radiation of Australasian songbirds. Mol. Phylogenet. Evol. 55: 1087-1102.

Groth, J. G., and G. F. Barrowclough. 1999. Basal divergences in birds and the phylogenetic utility of the nuclear RAG-1 gene. Mol. Phylogenet. Evol. 12:115-123.

Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52:696-704.

Hackett, S. J., R. T. Kimball, S. Reddy, R. C. K. Bowie, E. L. Braun, M. J. Braun, J. L. Chojnowski et al. 2008. A phylogenomic study of birds reveals their evolutionary history. Science 320:1763-1768.

Hillis, D. M., T. A. Heath, and K. John. 2005. Analysis and visualization of tree space. Syst. Biol. 54:471-482.

Jennings, W. B., and S. V. Edwards. 2005. Speciational history of Australian grass finches (*Poephila*) inferred from thirty gene trees. Evolution 59:2033-2047.

Johnson, K. P. 2004. Deletion bias in avian introns over evolutionary timescales. Molecular Biology and Evolution 21:599-602.

Joseph, L., and K. E. Omland. 2009. Phylogeography: its development and impact in Australo-Papuan ornithology with special reference to paraphyly in Australian birds. Emu 109:1-23.

Joseph, L., G. J. Adcock, C. Linde, K. E. Omland, R. Heinsohn, R. T. Chesser, and D. Roshier.

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

2009. A tangled tale of two teal: population history of the grey *Anas gracilis* and chestnut teal *A. castanea* of Australia. J.Avian Biol. 40:430-439.

Karl, S. A., and J. C. Avise. 1993. PCR-based assays of mendelian polymorphisms from anonymous single-copy nuclear DNA: techniques and applications for population genetics. Mol. Biol. Evol. 10:342-361.

Keast, A. 1961. Bird speciation on the Australian continent. Bulletin of the Museum of Comparative Zoology, Harvard College 123:305-495.

Kishino, and Hasegawa. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order of the Hominoidea. J. Molec. Evol. 29:170-179.

Knowles, L. L., and B. C. Carstens. 2007. Delimiting species without monophyletic gene trees. Systematic Biology 56:887-895.

Knowles, L. L. 2009. Statistical phylogeography. Ann. Rev. Ecol. Evol. Syst. 40:593-612.

Kubatko, L. S., and J. H. Degnan. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56:17-24.

Lake, J. A., J. A. Servin, C. W. Herbold, and R. G. Skophammer. 2008. Evidence for a New Root of the Tree of Life. Syst. Biol. 57:835-843.

Leaché, A. D. 2009. Species tree discordance traces to phylogeographic clade boundaries in North American fence lizards (*Sceloporus*). Syst. Biol. 58:547-559.

Leaché, A. D., and B. Rannala. 2011. The Accuracy of Species Tree Estimation under Simulation: A Comparison of Methods. Systematic Biology 60:126-137.

Lee, J. Y., and S. V. Edwards. 2008. Divergence across Australia's Carpentarian Barrier: statistical phylogeography of the red-backed fairy-wren (*Malurus melanocephalus*).

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

Evolution 62:3117-3134.

Leigh, J. W., E. Susko, M. Baumgartner, and A. J. Roger. 2008. Testing congruence in phylogenomic analysis. Syst. Biol. 57:104-115.

Librado, P., and J. Rozas. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451-1452.

Linnen, C. R., and B. D. Farrell. 2008. Comparison of methods for species-tree inference in the sawfly genus *Neodiprion* (Hymenoptera: Diprionidae). Syst. Biol. 57:876-890.

Liu, L. 2008. BEST: Bayesian estimation of species trees under the coalescent model. Bioinformatics 24:2542-2543.

Liu, L., and D. K. Pearl. 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. Syst. Biol. 56:504-514.

Liu, L., D. K. Pearl, R. T. Brumfield, and S. V. Edwards. 2008. Estimating species trees using multiple-allele DNA sequence data. Evolution 62:2080-2091.

Liu, L., and S. V. Edwards. 2009. Phylogenetic analysis in the anomaly zone. Syst. Biol. 58:452-460.

Liu, L., L. Yu, D. K. Pearl, and S. V. Edwards. 2009. Estimating species phylogenies using coalescence times among sequences. Syst. Biol. 58:468-477.

Liu, L., and L. Yu. 2010. Phybase: an R package for species tree analysis. Bioinformatics 26:962-963.

Maddison, D. R., and W. P. Maddison. 2000. MacClade version 4: Analysis of phylogeny and character evolution. Sinauer Associates, Sunderland, Massachusetts.

Maddison, W. P. 1997. Gene trees in species trees. Syst. Biol. 46:523-536.

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

Maddison, W. P., and L. L. Knowles. 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. 55:21-30.

Matthee, C. A., G. Eick, S. Willows-Munro, C. Montgelard, A. T. Pardini, and T. J. Robinson. 2007. Indel evolution of mammalian introns and the utility of non-coding nuclear markers in eutherian phylogenetics. Mol. Phylogenet. Evol. 42:827-837.

McCormack, J. E., H. T. Huang, and L. L. Knowles. 2009. Maximum Likelihood Estimates of Species Trees: How Accuracy of Phylogenetic Inference Depends upon the Divergence History and Sampling Design. Syst. Biol. 58:501-508.

Mulder, R. A., and A. Cockburn. 1993. Sperm competition and the reproductive anatomy of male superb fairy-wrens. Auk 110:588-593.

Murphy, W. J., T. H. Pringle, T. A. Crider, M. S. Springer, and W. Miller. 2007. Using genomic data to unravel the root of the placental mammal phylogeny. Genome Research 17:413-421.

Nylander, J. A. A. 2004. MrModeltest 2.0. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Peters, J. L., Y. Zhuravlev, I. Fefelov, A. Logie, and K. E. Omland. 2007a. Nuclear loci and coalescent methods support ancient hybridization as cause of mitochondrial paraphyly between gadwall and falcated duck (*Anas* spp.). Evolution 61:1992-2006.

Peters, J. L., and K. E. Omland. 2007b. Population structure and mitochondrial polyphyly in North American Gadwalls (*Anas strepera*). Auk 124:444-462.

Pollard, D. A., V. N. Iyer, A. M. Moses, and M. B. Eisen. 2006. Widespread discordance of gene trees with species tree in *Drosophila*: Evidence for incomplete lineage sorting. PLoS Genet. 2:1634-1647.

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

Robinson, D. F., and L. R. Foulds. 1981. Comparison of phylogenetic trees. Math. Biosci. 53:131-147.

Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

Rowe, M., M. R. Bakst, and S. Pruett-Jones. 2008. Good vibrations? Structure and function of the cloacal tip of male Australian Maluridae. J. Av. Biol. 39:348-354.

Rowe, M., and S. Pruett-Jones. 2008. Reproductive anatomy of male southern emu-wrens (*Stipiturus malachurus*) and striated grasswrens (*Amytornis striatus*). Emu 108:68-73.

Rowley, I., and E. Russell. 1997. Fairy-wrens and Grasswrens. Oxford University Press Inc., New York.

Schodde, R. 1982. The Fairy-wrens: a monograph of the Maluridae. Lansdowne Editions, Melbourne, Victoria.

Seo, T. K. 2008. Calculating bootstrap probabilities of phylogeny using multilocus sequence data. Mol. Biol. Evol. 25:960-971.

Shedlock, A. M., K. Takahashi, and N. Okada. 2004. SINEs of speciation: tracking lineages with retroposons. Trends in Ecology & Evolution 19: 545-553.

Shimodaira, H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51:492-508.

Shimodaira H. and M. Hasegawa (2001) CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics, 17, 1246-1247.

Stephens, M., N. J. Smith, and P. Donnelly. 2001. A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68:978-989.

Swaddle, J. P., J. Karubian, and S. Pruett-Jones. 1999. A novel evolutionary pattern of reversed

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

sexual dimorphism in fairy-wrens: implications for sexual selection. Behav. Ecol. 11:345-349.

Swofford, D. L. 2003. PAUP*. Phylogenetic analysis using parsimony (*and other methods) version 4. Sinauer Associates, Sunderland, Massachusetts.

Takahata, N. 1989. Gene genealogy in three related populations: consistency probability between gene and population trees. Genetics 122:957-966.

Than, C., D. Ruths, and L. Nakhleh. 2008. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. BMC Bioinformatics 9: 322.

Than, C., and L. Nakhleh. 2009. Species Tree Inference by Minimizing Deep Coalescences. PloS Computational Biology 5: e1000501.

Thomson, R. C., A. M. Shedlock, S. V. Edwards, and H. B. Shaffer. 2008. Developing markers for multilocus phylogenetics in non-model organisms: A test case with turtles. Mol. Phylogenet. Evol. 49:514-525.

Townsend, T. M., R. E. Alegre, S. T. Kelley, J. J. Wiens, and T. W. Reeder. 2008. Rapid development of multiple nuclear loci for phylogenetic analysis using genomic resources: an example from squamate reptiles. Mol. Phylogenet. Evol. 47:129-142.

Webster, M. S., C. W. Varian, and J. Karubian. 2008. Plumage color and reproduction in the red-backed fairy-wren: Why be a dull breeder? Behav. Ecol. 19:517-524.

Zink, R. M., and G. F. Barrowclough. 2008. Mitochondrial DNA under siege in avian phylogeography. Molecular Ecology 17:2107-2121.

Zwickl, D. J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. The University of

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

Texas, Austin.

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1. Descriptive statistics for the 18 loci used in this study. Sequence length includes alignment gaps. The substitution model was estimated using MrModelTest and the number of interspecific coalescent events was counted in MacClade. See methods for details.

Locus	No. chromosomes	Length (bp)	Variable sites (%)	Parsimony informative sites (%)	No. indels	GC%	Substitution model	Min. no. interspecific coalesent events	Observed no. interspecific coalescent events
Anonymous loci									
AL06	98	474	71 (15.0)	60 (12.7)	8	41.8	GTR+G	24	27
AL16	98	393	121 (30.8)	99 (25.2)	11	44.6	HKY+I	24	29
AL23	98	439	115 (26.2)	103 (23.5)	15	41.4	GTR+G	24	29
AL26	98	404	125 (30.9)	105 (26.0)	8	51.4	HKY+G	24	27
AL28	98	460	129 (28.0)	107 (23.3)	6	44.4	GTR+G	24	26
Mean		**433.8**	**112.2 (25.9)**	**90.0 (20.5)**	**9.6**	**44.7**		**24**	**27.6**
Exons									
FSHR	86	372	60 (16.1)	47 (12.6)	0	46.9	K80+G	24	31
MEK1	86	397	26 (6.5)	18 (4.5)	3	47.5	HKY+G	24	31
PTPN12	86	812	97 (11.9)	78 (9.6)	0	46.8	HKY+I+G	24	26
TEX10	86	545	52 (9.5)	40 (7.3)	0	42.9	HKY+I	24	35
TNNT3	96	297	43 (14.5)	33 (11.1)	1	52.8	HKY+G	24	28
TRAF6	86	522	59 (11.3)	43 (8.2)	0	47.4	HKY+I	24	28
UBN1	86	328	60 (18.3)	50 (15.2)	9	42.7	GTR+G	24	28
Mean		**467.7**	**56.7 (12.1)**	**43.6 (9.3)**	**1.9**	**46.7**		**24**	**29.6**
Introns									
AB4	98	203	42 (20.7)	39 (19.2)	1	54.1	SYM+G	24	29
CDC132	86	424	145 (34.2)	125 (29.5)	12	37.6	HKY+G	24	28
HMGB2	90	501	116 (23.2)	105 (19.6)	16	38.1	GTR+G	24	25
RI2	98	257	54 (21.0)	49 (19.1)	6	60	GTR+G	24	29
TGFβ2	98	552	147 (26.6)	128 (23.2)	10	44.2	GTR+G	24	31
Mean		**384.0**	**100.8 (26.2)**	**80.0 (20.8)**	**9**	**46.8**		**24**	**28.4**
MtDNA									
ND2	57	447	207 207(46.3)	207(46.3)	0	47.5		24	24

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 2. Comparison among trees built from concatenated data (Phyml, ML, mrBayes) and from species tree methods under a concatenation framework using the approximately unbiased (AU) test (Shimodaira 2002) and Kishino-Hasegawa (1989) (KH) test .

Tree	Data set[c]	ΔLnL	AU	KH
Phyml	single[c]	MLE	0.627	0.544
ML (Paup)	single	0.9	0.633	0.456
mrBayes	single	3.7	0.455	0.356
STAR[a]	multi	4.7	0.496	0.364
STAR[a]	single	7.7	0.373	0.316
BEST	single	11.1	0.286	0.275
Bucky	single	11.2	0.255	0.223
Bucky	multi	17.5	0.132	0.153
MDC[a]	multi	67.8	0.001	0.002
MDC[b]	multi	74.5	0.001	0.001
MDC[b]	single	93.9	3e-04	0.001

[a]Bootstrap consensus tree (see Methods)

[b]Single heuristic search

[c]Single- or multi-allele data sets

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 1. Variation among loci. The proportion of variable sites and parsimony informative sites per locus are given in the white and black bars, respectively. All sample sizes refer to the full data set for each locus.

Figure 2. Examples of gene trees with the single and multiple allele data sets. The gene for each tree is given at the upper left. In each tree the three major groups in the Maluridae are given in different colors: *Malurus* (black), basal malurids (white with black outline), emu-wrens (light gray) and grasswrens (dark gray). Branch lengths are proportional to substitutions per site within each tree, but not among trees. All gene trees were made using mrBayes. a-c represent gene trees made from the full data set and trees in d-f are made from the single-allele (single-allele) data set. In the multiple allele data sets (a-c) the number of additional parsimony steps over the minimum number of interspecific coalescent events in each tree is given in parentheses (see also Table 1). In trees a-c, the black boxes indicate areas of conspicuous incomplete lineage sorting that are detailed in Fig. 3.

Figure 3. Details of gene tree topologies for various clades (a-e). The four gene trees with numbers (a, b, d and e) correspond to boxes 1-4 in Fig. 2 (UBN1 did not appear in Fig. 2). Numbers beside species names designate individuals (see Appendix Table 1); lower case 'a' or 'b' indicate designations of alleles estimated using Phase (see Methods). Topologies only are given; branches are not proportional to lengths. Each branch is colored by species; each taxon name consists of a species, an individual number, and a phased allele (a or b). In fact many of the branches depicted are extremely short, some probably not significantly different from zero. Nonetheless for each gene depicted the cumulative effect of many such branches is enough for

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

the AU test to reject tree topologies in which all alleles of each species form monophyletic groups.

Figure 4. Visualization of gene tree dispersion in two-dimensional tree space using the matrix of Robinson-Foulds distances. Gene trees as estimated by mrBayes (a), Phyml (b) or BEST (c) were placed so that the distortion between the true multidimensional distance between pairs of trees and the distance in 2D space minimized. In panels a-c, Eighteen gene trees include seven exons ('E'), five introns ('I'), five anonymous loci ('A') and one mitochondrial ND2 gene ('M'), as follows: A_1) AL06, A_2) AL16, A_3) AL23, A_4) AL26, A_5) AL28, E_1) FSHR, E_2) MEK1, E_3) PTPN12, E_4) TEX10, E_5) TNNT3, E_6) TRAF6, E_7) UBN1, I_1) AB4, I_2) CDC132, I_3) HMG-2, I_4) RI2, I_5) TGFb2, M) ND2. In addition, five multilocus species trees are depicted in panels a-c, corresponding to BEST (B), STAR (S), BUCKy (Y), MDC (DC) and concatenation (C). Shading around groups of gene trees indicate those genes whose phylogenetic signal is not significantly discordant as measured by Concaterpillar on the single-allele data set. In panel d, the distribution of Robinson-Foulds distances among the 18 gene trees is given for mrBayes, Phyml and BEST.

Figure 5. Species tree reconstructed using the Bayesian Estimation of Species Tree (BEST) program on the single-allele data set, which was a complete matrix containing 26 alleles across the 25 species and outgroup. Numbers on branches indicate posterior probabilities. Species names with asterisks indicate those depicted in figures to the right, in order from top to bottom. Bird figures by Peter Marsack as in Rowley and Russell (1997). See text for details.

Figure 6. Species trees reconstructed using the (a) concatenation method, (b) STAR and (c)

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

BUCKy. Single-allele (25 ingroup sequences) and multiple-allele (86 sequences across 25 species and outgroup) data sets were used for the trees on the left and on the right, respectively, of panels b and c. Numbers on nodes represent posterior probabilities (a), bootstrap percentages (b,c) and concordance percentages (d,e). In the multiple-allele BUCKy tree (c, right), the summary tree placed alleles within each species as monophyletic groups. We therefore collapsed these clusters and represented them as a single tip.

Figure 7. Phylogeny reconstructed using unordered parsimony on 102 independent indels, single-allele data set, with 27 invariant indels. This tree is a consensus of 120 equally parsimonious trees of length 95. The numbers above branches indicate the number of unambiguously reconstructed deletions (-) and insertions (+). The asterisks denote character changes that are arbitrarily assigned to one or the other side of the root due to lack of an outgroup and whose direction is ambiguous.

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

Appendix Table 1

No	Scientific name	Study number[d]	Common name	Depository[b]	Specimen No.	State/Country[c]
1[d,e]	*Amytornis ballarae*	1	Kalkadoon Grasswren	ANWC	41740	QLD
2[e]	*Amytornis ballarae*	2	Kalkadoon Grasswren	ANWC	41741	QLD
3[e]	*Amytornis ballarae*	3	Kalkadoon Grasswren	ANWC	41742	QLD
4[e]	*Amytornis barbatus*	1	Grey Grasswren	ANWC	40054	QLD
5	*Amytornis barbatus*	2	Grey Grasswren	ANWC	40141	SA
6[d,e]	*Amytornis barbatus*	3	Grey Grasswren	ANWC	41788	QLD
7	*Amytornis dorotheae*	1	Carpentarian Grasswren	ANWC	28406	QLD
8[d,e]	*Amytornis dorotheae*	2	Carpentarian Grasswren	ANWC	28407	QLD
9[d,e]	*Amytornis goyderi*	1	Eyrean Grasswren	ANWC	40086	SA
10[e]	*Amytornis goyderi*	2	Eyrean Grasswren	ANWC	40138	SA
11[e]	*Amytornis goyderi*	3	Eyrean Grasswren	ANWC	40169	SA
12[d,e]	*Amytornis housei*	1	Black Grasswren	ANWC	24307	WA
13[e]	*Amytornis housei*	2	Black Grasswren	ANWC	39173	WA
14[e]	*Amytornis housei*	3	Black Grasswren	ANWC	39174	WA
15[d,e]	*Amytornis merrotsyi*	1	Short-tailed Grasswren	ANWC	28202	SA
16[e]	*Amytornis merrotsyi*	2	Short-tailed Grasswren	ANWC	40615	SA
17	*Amytornis merrotsyi*	3	Short-tailed Grasswren	ANWC	48272	SA
18[d,e]	*Amytornis purnelli*	1	Dusky Grasswren	ANWC	40221	NT
19[e]	*Amytornis purnelli*	2	Dusky Grasswren	ANWC	40239	NT
20[e]	*Amytornis purnelli*	3	Dusky Grasswren	ANWC	40258	NT
21[d,e]	*Amytornis textiles*	1	Thick-billed Grasswren	ANWC	40176	SA
22[e]	*Amytornis textiles*	2	Thick-billed Grasswren	ANWC	40192	SA
23	*Amytornis textiles*	3	Thick-billed Grasswren	ANWC	40314	SA
24[d,e]	*Amytornis striatus*	1	Striated Grasswren	ANWC	31651	NSW
25[e]	*Amytornis striatus*	2	Striated Grasswren	ANWC	40614	SA
26[e]	*Amytornis striatus*	3	Striated Grasswren	ANWC	48515	QLD
27[d,e]	*Stipiturus malachurus*	1	Southern Emu-wren	ANWC	20748	SA

http://mc.manuscriptcentral.com/systbiol

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

28[ee]	*Stipiturus malachurus*	2	Southern Emu-wren	ANWC	31732	WA
29[e]	*Stipiturus malachurus*	3	Southern Emu-wren	ANWC	45971	TAS
30[d,e]	*Stipiturus mallee*	1	Mallee Emu-wren	ANWC	40418	SA
31[d,e]	*Stipiturus ruficeps*	1	Rufous-crowned Emu-wren	ANWC	39914	QLD
32[ee]	*Stipiturus ruficeps*	2	Rufous-crowned Emu-wren	ANWC	48922	NT
33[d,e]	*Clytomias insignis*	1	Orange-crowned Fairy wren	KU	7911	PNG
34[d,e]	*Malurus alboscapulatus*	1	White-shouldered Fairy wren	KU	12171	PNG
35[e]	*Malurus alboscapulatus*	2	White-shouldered Fairy wren	KU	12172	PNG
36[d,e]	*Malurus cyanocephalus*	1	Emperor Fairy wren	KU	7564	PNG
37	*Malurus cyanocephalus*	2	Emperor Fairy wren	KU	7565	PNG
38[d,e]	*Malurus grayi*	1	Broad-billed Fairy wren	KU	7082	PNG
39[e]	*Malurus grayi*	2	Broad-billed Fairy wren	KU	7083	PNG
40[d,e]	*Malurus coronatus*	1	Purple-crowned Fairy wren	BMNHC	60807	NT
41[d,e]	*Malurus amabilis*	1	Lovely Fairy wren	ANWC	31303	QLD
42[ee]	*Malurus amabilis*	2	Lovely Fairy wren	ANWC	43015	QLD
43	*Malurus amabilis*	3	Lovely Fairy wren	ANWC	51782	QLD
44[ee]	*Malurus cyaneus*	1	Superb Fairy wren	ANWC	34555	ACT
45[d,e]	*Malurus cyaneus*	2	Superb Fairy wren	ANWC	46115	VIC
46[ee]	*Malurus cyaneus*	3	Superb Fairy wren	ANWC	46669	SA
47[d,e]	*Malurus elegans*	1	Red-winged Fairy wren	ANWC	31938	WA
48[d,e]	*Malurus pulcherrimus*	1	Blue-breasted Fairy wren	ANWC	28233	SA
49[f]	*Malurus lamberti*[f] *(M. pulcherrimus)*	3 (2)	Blue-breasted Fairy wren	ANWC	33803	WA
50[ee]	*Malurus pulcherrimus*	2	Blue-breasted Fairy wren	ANWC	46759	SA
51[d,e]	*Malurus splendens*	1	Splendid Fairy wren	ANWC	28009	QLD
52[ee]	*Malurus splendens*	2	Splendid Fairy wren	ANWC	44501	NSW
53[d,e]	*Malurus lamberti*	1	Variegated Fairy wren	ANWC	31655	NSW
54[ee]	*Malurus lamberti*	2	Variegated Fairy wren	ANWC	46475	SA

http://mc.manuscriptcentral.com/systbiol

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

55[d,e]	*Malurus melanocephalus*	1	Red-backed Fairy wren	ANWC	29906	QLD
56[e]	*Malurus melanocephalus*	2	Red-backed Fairy wren	ANWC	49079	QLD
57[d,e]	*Malurus leucopterus*	1	White-winged Fairy wren	ANWC	20947	QLD
58[e]	*Malurus leucopterus*	2	White-winged Fairy wren	ANWC	33099	WA
59	*Greygone olivacea*	n/a	White-throated Greygone	MCZ	336023	NSW

All 59 individuals were used to generate individual gene trees (data set 1).

[a]Indicates the number assigned to the individual in Fig. 4.

[b] Depository Abbreviations: ANWC, the Australian National Wildlife Collection; BMNHC, the Burke Museum of Natural History and Culture; KU, the KU Natural History Museum; MCZ, the Museum of Comparative Zoology at Harvard University.

[c]Qld. = Queensland, SA= South Australia, WA= Western Australia, TAS=Tasmania, NT=Northern Territory, ACT=Australian Capital Territory, NSW=New South Wales, VIC=Victoria, PNG=Papua New Guinea.

[d,]indicates the 26 individuals from which alleles were chosen for the single-allele analyses (data set 3).

[e]indicates the individuals from which both alleles were chosen for the full analyses (data set 2).

[f]Indicates the *M. lamberti* specimen that was at first mistaken for a *M. pulcherrimus* specimen (see text).

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

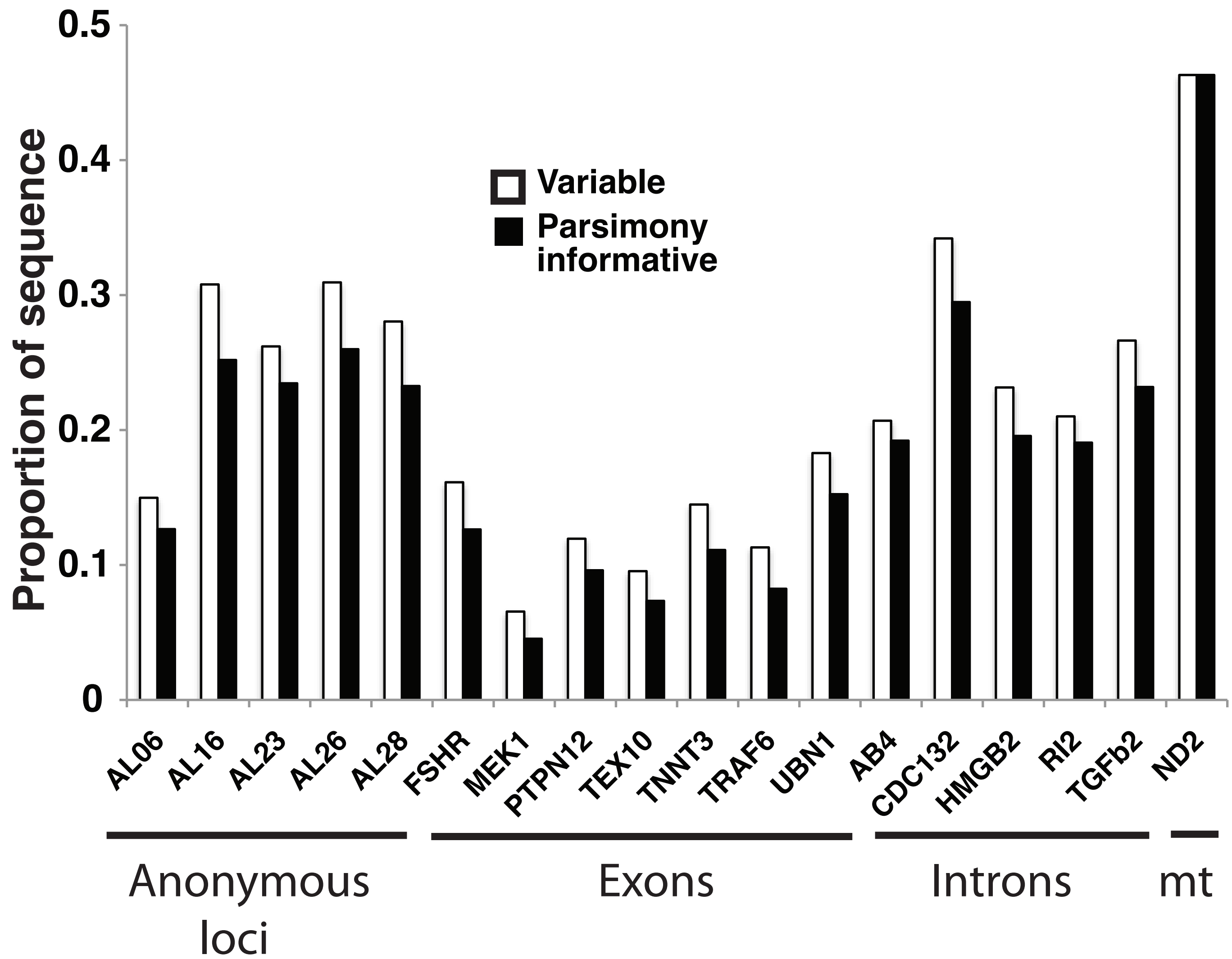
Appendix Table 2

	Locus	Forward	Reverse	Reference
ANONYMOUS LOCI	Mame-AL06	AGAAGAATCCGTGTGCCAAC	ATGTTCAGCACAACCACAGC	Lee and Edwards 2008
	Mame-AL16	GCAGGGAGGTGTGATTATAGC	AGCCCAAAGTTGTCAGAAGC	Lee and Edwards 2008
	Mame-AL23	TGCATTCACACCAGGAATTG	GTGCTGGCACTGAAACTTCC	Lee and Edwards 2008
	Mame-AL26	ATGCCAGCTGCAAAGGTTAC	ATGGGCAGTTGTTTGCTTTC	Lee and Edwards 2008
	Mame-AL28	AGAGCCAGGAAAACCTCTTT	TGGAGGTGATTGAATGAATG	Lee and Edwards 2008
INTRONS	AB4	WCCTCTKGCAGGAACAAA	GTCARCACCATCTTTCTTGTACT	Waltari and Edwards 2002
	RI2	TGGTGGTCTGYAAGCCCAT	TGGACCACGAACATGTAGAT	Waltari and Edwards 2002
	TGFb2	TTGTTACCCTCCTACAGACTTGAGTC	GACGCAGGCAGCAATTATCC	Sorenson *et al.* 2004
	CDC132	TCTGGGAACAGATCTGTC	AAACTTCAGACTTACTGCC	Backström *et al.* 2007
	HMG-2	GAAATGTGGTCTGAACAGTC	TTGCTCTTGGCACGATATGC	Backström *et al.* 2007
EXONS	FSHR	GCAGCTGAAGATTWYATTTC	TYTGCCAGTCTATGGCRT	From this study
	MEK1	CACCTCTTAAGGAAGAATGTCTTKA	GAGCTCTTRCCTTCTTYACCT	From this study
	PTPN12	AGTTGCCTTGTWGAAGGRGATGC	CTRGCAATKGACATYGGYAATAC	Townsend *et al.* 2008
	TEX10	GCAGCWGTGTTTACAGACAA	GYTSATCTTTGGTCGTAAGC	From this study
	TNNT3	AGCAAGAAGGCAGGAGCC	CCAGRGATTTGTACACAGCAATCTA	From this study
	TRAF6	ATGCAGAGGAATGARYTGGCACG	AGGTGGCTGTCRTAYTCYCCYYGC	Townsend *et al.* 2008
	UBN1	GTACCTCCAAGYCCAGTTACC	AAAAGCATCCCTTTGGAG	From this study
mtDNA	ND2	GGCCCATACCCCGRAAATG	RGAKGAGAARGCYAGGATYTTKCG	Sorenson *et al.* 1999

All genes identified by the official symbol of their respective human homologs in NCBI's Entrez Gene: AB4, aldolase B intron 4; RI2, rhodopsin intron2; TGFß2, transforming growth factor-ß2 intron 5; CDC132, coiled-coil domain containing 132; HMG-2, high mobility group protein B2; FSHR, follicle stimulating hormone receptor; MEK1, MAP kinase-kinase; PTPN12, protein tyrosine phosphatase non-receptor type 12; TEX10, testis expressed 10; TNNT3, troponin T type 3; TRAF6, TNF receptor-associated factor 6; UBN1, ubinuclein 1; ND2, NADH dehydrogenase subunit 2.

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

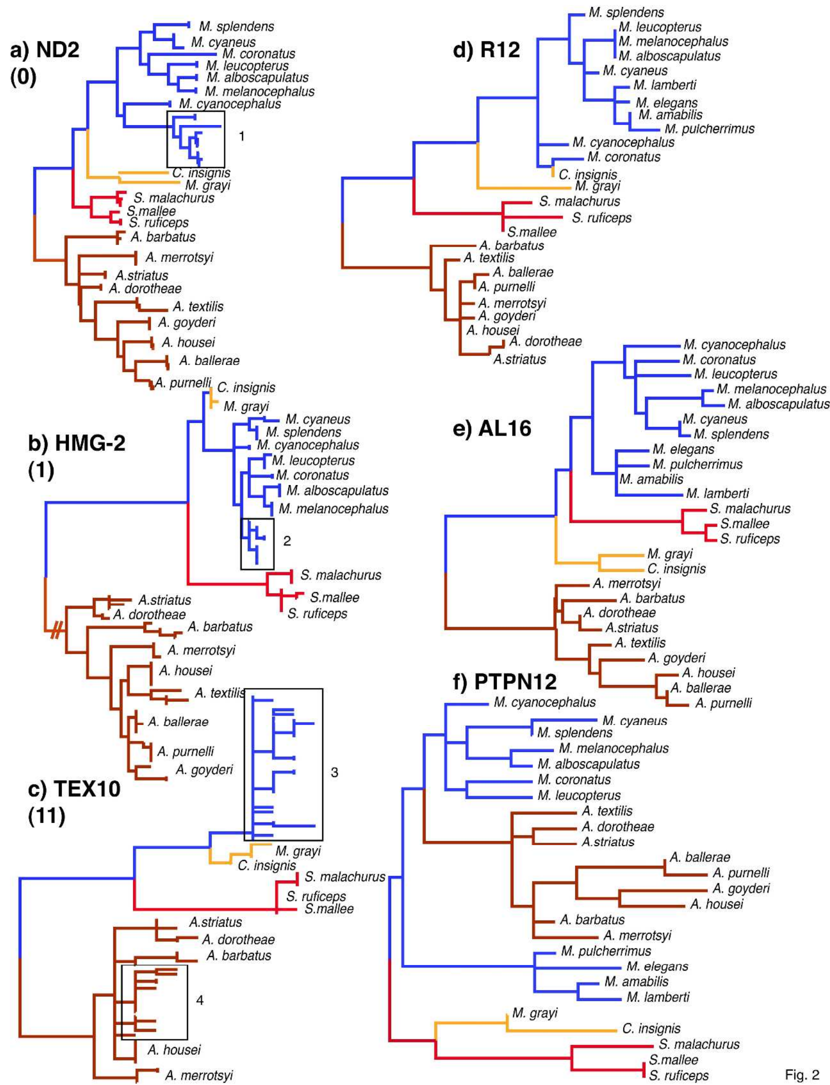
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

Figure 2

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



308x426mm (300 x 300 DPI)

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

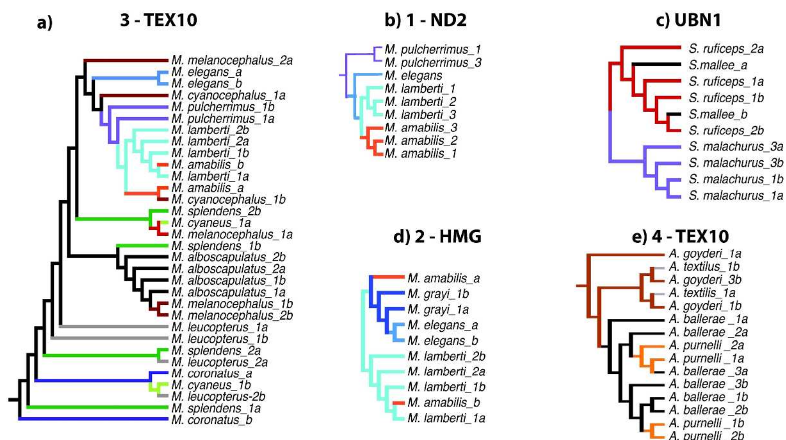


Fig. 3

264x388mm (300 x 300 DPI)

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

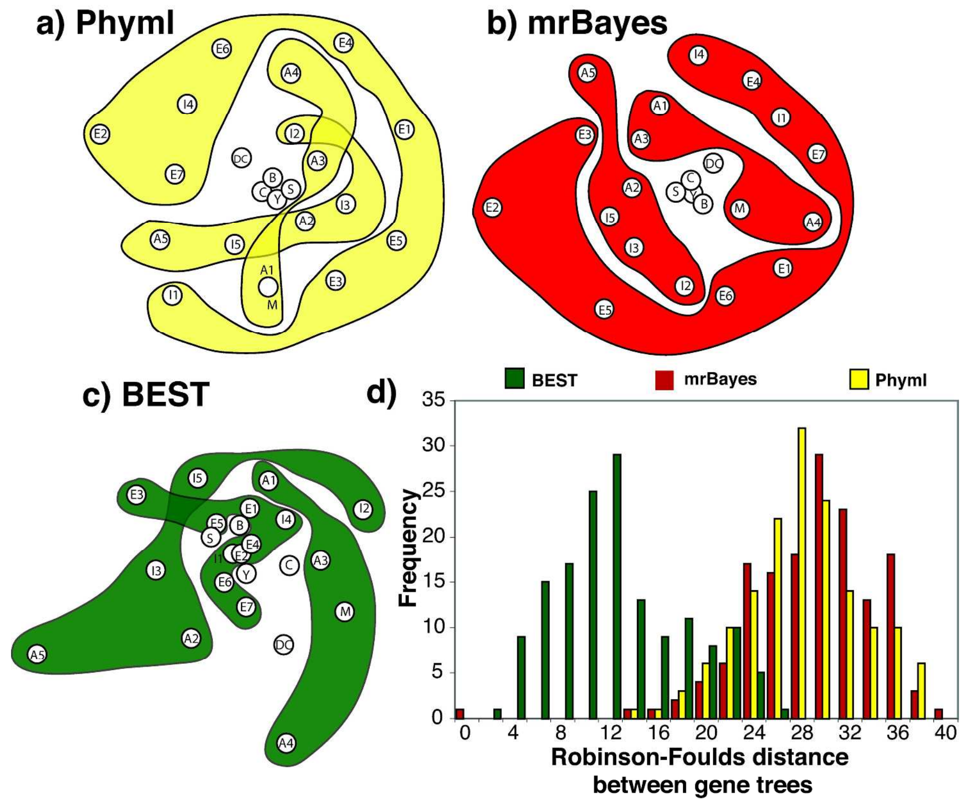
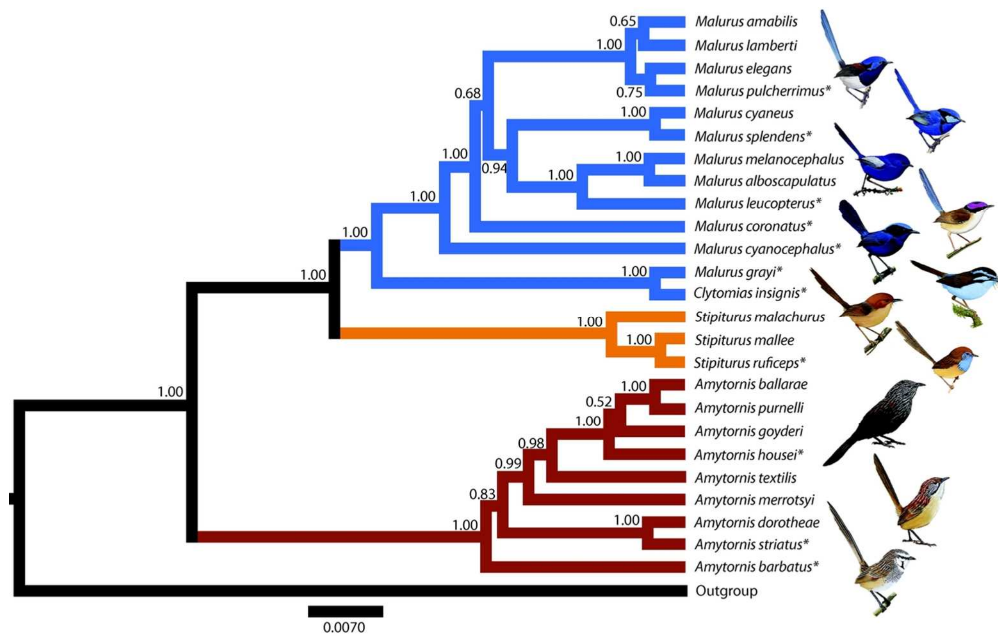


Fig. 4

263x334mm (300 x 300 DPI)

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



119x76mm (300 x 300 DPI)

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

a) mrBayes, concatenation

b) MDC

c) STAR, single allele

d) STAR, multiple allele

e) BUCKY, single allele

f) BUCKY, multiple allele

Fig. 6

270x381mm (300 x 300 DPI)

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016

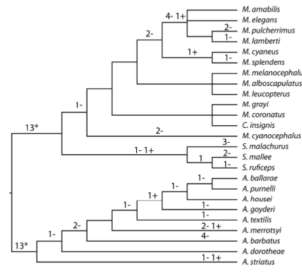


Fig. 7

235x91mm (300 x 300 DPI)

Downloaded from http://sysbio.oxfordjournals.org/ by guest on September 22, 2016