Molecular Phylogenetics and Evolution 58 (2011) 492–501

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Phylogenetic relationships in the spoon tarsus subgroup of Hawaiian *drosophila*: Conflict and concordance between gene trees

Richard T. Lapoint *, Alexander Gidaya, Patrick M. O'Grady

University of California, Department of Environmental Science, Policy and Management, 137 Mulford Hall, Berkeley, CA 94720, United States

ARTICLE INFO

Article history:
Received 21 August 2010
Revised 24 December 2010
Accepted 24 December 2010
Available online 30 December 2010

Keywords:
Hawaiian *Drosophila*
Spoon tarsus subgroup
Species tree

ABSTRACT

The Hawaiian Drosophilidae contains approximately 1000 species, placed in species groups and subgroups based largely on secondary sexual modifications to wings, forelegs and mouthparts. Members of the spoon tarsus subgroup possess a cup-shaped structure on the foretarsi of males. Eight of the twelve species in this subgroup are found only on the Big Island of Hawaii, suggesting that they have diverged within the past 600,000 years. This rapid diversification has made determining the relationships within this group difficult to infer. We use 13 genes, including nine rapidly evolving nuclear loci, to estimate relationships within the spoon tarsus species, as well as to test the monophyly of this subgroup. A variety of analytical approaches are used, including individual and concatenated analyses, Bayesian estimation of species trees and Bayesian untangling of concordance knots. We find widespread agreement between phylogenetic estimates derived from different methods, although some incongruence is present. Notably, our analyses suggest that the spoon tarsus subgroup, as currently defined, is not monophyletic.

1. Introduction

The Hawaiian Islands, located approximately 4000 km from the nearest continent, are the most isolated island chain in the world (Carson and Clague, 1995). These volcanic islands rise out of the Pacific where lava seeps through a "hot spot" in the Earth's crust, and over time becomes large enough to sustain life. As the Pacific Plate moves, islands are carried northwest. When an island moves off the hotspot, it become dormant and gradually erodes back into the sea (Carson and Clague, 1995). This leads to a datable progression with young islands in the southwest and sequentially older islands in the northwest (see Fig. 1). High mountains that gradually slope into the sea characterize the young islands. Older islands are shorter, steeper and deeply carved by the action of wind and water (Price and Clague, 2002).

The Hawaiian chain is currently composed of eight high islands, those that are high enough to catch the moist trade winds and produce rainforest habitats. The high percentage of species found on the Hawaiian Islands has made it home to some of the highest levels of endemism in the United States (Eldredge and Evenhuis, 2003). Price and Clague (2002) have recently reviewed several lineages endemic to Hawaii that have originated from a single or few initial colonizers. In each of the lineages of endemic Hawaiian species that were reviewed it appears that the colonization occurred after the formation of the current high islands (Baldwin and Sanderson, 1998; Fleischer et al., 1998). However a few radiations are estimated to have colonized the island chain well before the formation of the current high islands, and have subsequently progressed down to the younger islands as they form (Givnish et al., 1996; Jordan et al., 2003).

The oldest inhabitants of the island chain are the Hawaiian Drosophilidae, a radiation derived from a single colonization event approximately 25 million years ago (Russo et al., 1995). Since this time they have diversified into a clade of 1000 species (O'Grady et al., 2010), occupying a wide variety of ecological niches (Heed, 1968; Montgomery, 1975; Magnacca et al., 2008) and displaying impressive morphological diversity (Hardy, 1965). They have diversified into many different niches, exploiting nearly 40% of the native plant families and even more bizarre substrates like spider eggs (Wirth, 1952). Most species also display marked sexual dimorphism, with males possessing elaborate secondary characters, such as wing patterning, elongate antennae, tusk-like mouthparts and elaborate processes on the tarsi (e.g., Stark and O'Grady, 2009), that they use in copulation. Mating displays are likewise diverse (Spieth, 1966). Morphological and molecular characters have been useful in defining several groups that have been tested by rigorous phylogenetic analysis (Baker and DeSalle, 1997; Bonacum, 2001; Bonacum et al., 2005; Carson and Stalker, 1969; Kambysellis et al., 1995; O'Grady and Zilversmit, 2004; O'Grady et al., in press).

Males in the spoon tarsus clade have an eponymous cup-like second tarsi on their forelegs used in positioning females during mating (Spieth, 1966; Stark and O'Grady, 2009). All species are

* Corresponding author.
E-mail address: rlapoint@berkeley.edu (R.T. Lapoint).

1055-7903/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2010.12.015

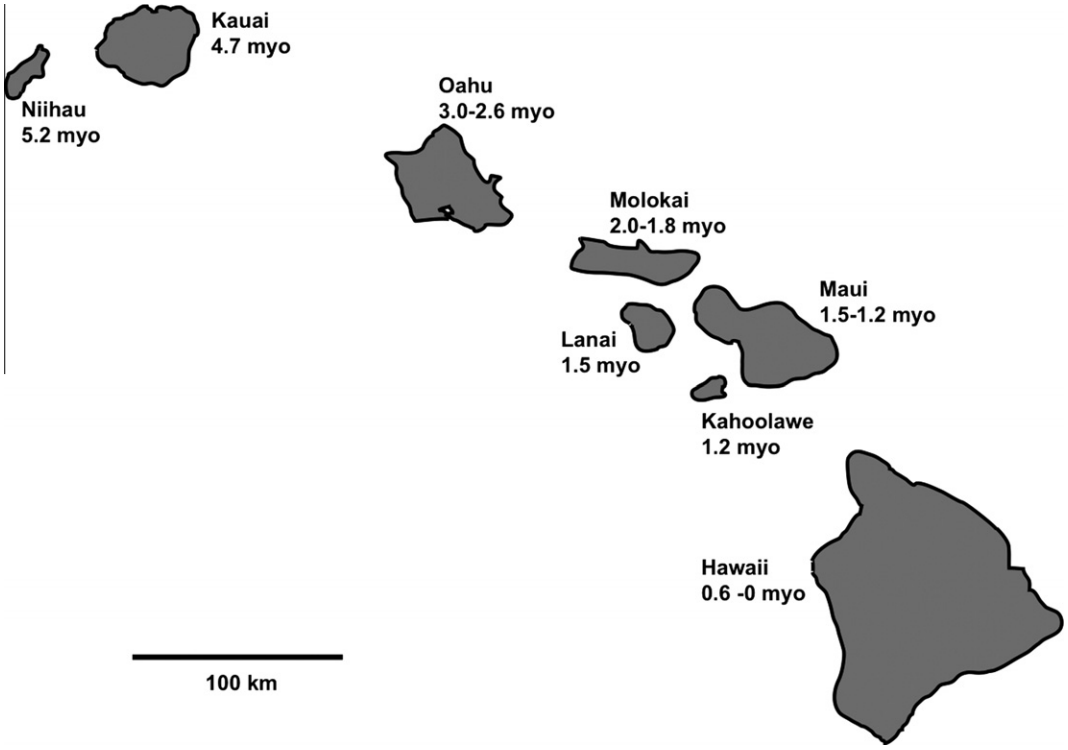

Fig. 1. Map of the Hawaiian Island Chain with the ages of each island based on K–Ar dating. The age of the islands increase from east to west.

ecologically similar, utilizing leaves of the endemic plant group *Araliaceae* as larval substrate (Magnacca et al., 2008) and are found in a similar range of habitats (Lapoint et al., 2009). Of the 12 species in this group there are eight present on the island of Hawaii (Big Island) seven of which are endemic (Lapoint et al., 2009). Since the Big Island is less than 600,000 years old these species are expected to have diverged recently and personal observations of density in the field suggest they exist in large population sizes, making them prone to the effects of deep coalescences.

Most recent radiations, where the coalescent is large due to either a recent time common ancestor or large population sizes, are known as difficult subjects for phylogenetic estimation (Belfiore et al., 2008; Degnan and Rosenberg, 2006). Resolving the species tree for such groups is especially important since these radiations can offer insight into many different aspects of biological processes that are better understood by placing the questions in a phylogenetic context. There are four main issues when estimating phylogenies for recent radiations: (1) few variable loci, (2) introgression, (3) cryptic species, (4) incomplete lineage sorting. First, there is a difficulty in finding markers variable enough to be informative at shallow divergences. With the advent of genomic resources (Drosophila 12 Genomes Consortium, 2007) the first issue has become less daunting since marker discovery methods have vastly improved (Edwards, 2008; Thomson et al., 2008; Wahlberg and Wheat, 2008). Introgression between species is another factor that can confound phylogenetic reconstruction. In the case of recent divergences, the expectation that all alleles follow a coinherited evolutionary history is not met. Mating barriers, for example, may not be complete and gene flow between incipient species may complicate phylogenetic estimation (Maddison, 1997; Leache and McGuire, 2006). Cryptic species are distinct evolutionary lineages that have not diverged morphologically and when included in phylogenies can make identified "species" paraphyletic. In recent radiations, lineages may become genetically distinct, but may not have had time to diverge morphologically and can warrant a reevaluation of species delimitation.

Finally, incomplete lineage sorting is a common and difficult issue to address in recent radiations (Degnan and Rosenberg, 2006). Incomplete lineage sorting is expected to be widespread in the case of recent radiations, and, by chance alone, genealogies that support the wrong topology can be more common than those that support the true topology (Rokas et al., 2003; Pollard et al., 2006). Clades with large population sizes (θ) and short divergence times are especially prone to this issue in simulation studies. In the case of Galapagos finches, silverswords and African cichlids, their rates of diversification and population sizes are within this range and they can be inferred to be prone to lineage sorting issues (McCormack et al., 2009). Empirical evidence shows that the issue of deep coalescences is real, as phylogenetic estimation within these groups has proven to be very difficult (Freeland and Boag, 1999; Lopez-Fernandez et al., 2010).

Several methods have been developed to address this issue of gene tree conflict. Concatenation is a total evidence method that combines all available information and analyzes it together as a single gene, assuming that the independent loci have evolved with a single evolutionary history that is predominantly displayed in the data (Degnan and Rosenberg, 2006). It has proven to be fairly robust in empirical studies in that this method generally finds well-resolved trees with good support (Belfiore et al., 2008; Thomson et al., 2008; Brumfield et al., 2008). A concatenated dataset can be partitioned into classes of data that are assumed to have evolved under the same model, such as gene or codon position (Ronquist and Huelsenbeck, 2003). In datasets composed of sequences evolving under different models, under parameterization can be mediated by partitioning and phylogenetic estimation improved (Brandley et al., 2005). Despite these benefits, drawbacks exist that make the results of this method questionable. Long branches separated by short internal nodes are especially prevalent when large amounts of data are used and can lead to positively misleading topologies (Gadagkar et al., 2005).

Another way of addressing the issue of lineage sorting is a suite of methods that utilize the coalescent. They attempt to find

the best species tree by estimating each gene genealogy independently and assuming that conflict between gene trees is due solely to incomplete lineage sorting. Under these approaches, the best estimate of phylogeny is the one that minimizes the number of deep coalescence events. Several algorithms (Liu et al., 2008; Kubatko, 2009) can perform this estimate in either a Bayesian or maximum likelihood framework. Although these methods are very parameter rich and can be computationally intensive (Knowles, 2009), they provide a good estimate of phylogeny when most of the conflict between gene trees is due to incomplete lineage sorting.

Here we reconstruct phylogenetic relationships in the spoon tarsus subgroup analyzing data in individual and partitioned concatenated matrices using both coalescent and concordance frameworks to compare their ability to find a resolved and robust phylogeny. Specifically, we are interested in testing recent taxonomic hypotheses about the composition of this group, to determine evolutionary relationships between closely related taxa, and ascertain whether the Big Island species cluster is comprised of discrete species or an irresolvable species cluster.

2. Materials and methods

2.1. Sampling

Eleven of the 12 spoon tarsus species were obtained for this analysis. *Drosophila mimiconformis*, a rare species endemic to the rainforests of Molokai, was not collected. We included multiple representatives from different volcanic mountains or different sides of the same mountain (Kona or Hilo) for species found on the Big Island in order to rigorously test the monophyly of these species. Multiple populations of *Drosophila waddingtoni* from various localities on Maui and Molokai were also included to test whether: (a) Maui Nui populations were ancestral to Big Island populations and follow the progression rule down the island chain (Wagner and Funk, 1995), or (b) if the Maui Nui populations were recent and resulted from back colonization from the Big Island (Wagner and Funk, 1995). We included *Drosophila grimshawi* (picture wing group), *Drosophila diamphidiopoda* (*antopocerus* group) and *Drosophila expansa* (bristle tarsus subgroup) as outgroups to test the monophyly of the spoon tarsus subgroup (Table 1). Both the *antopocerus* group and bristle tarsus subgroup are part of the same species clade as the spoon tarsus subgroup, the AMC (*antopocerus*, modified tarsus and ciliated tarsus clade). *D. grimshawi* is from a distantly related clade (O'Grady et al., in press) and was used as a known outgroup for all analyses.

Collections were made by sweeping leaf litter and aspirating specimens directly from sponges baited with fermenting banana. Specimens were stored in 95% EtOH for identification and DNA extraction at UC, Berkeley. Species identifications were performed by the authors using the key provided in Lapoint et al. (2009). Genomic DNA was extracted from individual flies using the Qiagen DNeasy DNA extraction kit (Qiagen, Inc). The only departure from this protocol was that individuals were soaked in Proteinase K instead of being macerated and subsequently preserved as point mounted vouchers. All voucher material has been deposited in either the B.P. Bishop Museum or the Essig Museum of Entomology at UC Berkeley. Table 1 lists the collection information and 6-digit collection code.

Table 1
Species sampling.

Species name	Barcode[b]	Taxonomic group[c]	Location[d]
atroscutellata[a]	070236a	ST	Kauai: Nualolo Trail, 3800'
conformis[a]	201312a	ST	Hawaii: HVNP, Olaa Tract
conformis	202493a	ST	Hawaii: Laupahoehoe, NARS, 4000'
conformis[a]	202486a	ST	Hawaii: Kau Forest Reserve, Hionamoa Stream
contorta[a]	200120a	ST	Maui: Waikamoi Forest Preserve, Pig Hunter's Trail
dasycnemia	202320a	ST	Hawaii: Saddle Road, Kipuka 9, Upper Waiakea FR
dasycnemia	202326b	ST	Hawaii: Puu O'umi NAR, Kohalas 5000ft
dasycnemia	202331a	ST	Hawaii: Laupahoehoe, NARS, 3700ft
dasycnemia[a]	200122a	ST	Hawaii: HVNP, Olaa Tract
fastigata[a]	070069a	ST	Oahu: Manoa Cliff Trail, 1800'
incognita[a]	202464a	ST	Hawaii: Ola'a Forest, pole 44, 3900'
incognita[a]	202333a	ST	Hawaii: Laupahoehoe, NARS, 3700ft
kikalaeleele[a]	202385a	ST	Hawaii: Kukui Opae, South Kona Forest Reserve, 3400ft
neutralis	202311a	ST	Hawaii: Stainback Highway, Tom's Trail, 3200'
neutralis	202319b	ST	Hawaii: Kau Forest Reserve, Hionamoa Stream
neutralis	202330a	ST	Hawaii: Laupahoehoe, NARS, 3700ft
neutralis[a]	202329b	ST	Hawaii: Kukui Opae, South Kona Forest Reserve, 3400ft
percnosoma	202394e	ST	Hawaii: Laupahoehoe, NARS, 3700ft
percnosoma	200125a	ST	Hawaii: HVNP, Olaa Tract
percnosoma[a]	202343d	ST	Hawaii: Kau Forest Reserve, Hionamoa Stream
sordidapex	202318a	ST	Hawaii: Kau Forest Reserve, Hionamoa Stream
sordidapex	202327a	ST	Hawaii: Kukui Opae, South Kona Forest Reserve, 3400ft
sordidapex	202332c	ST	Hawaii: Laupahoehoe, NARS, 3700ft
sordidapex[a]	202321a	ST	Hawaii: Ola'a Forest, Small tract, Transect 1
waddingtoni[a]	202431c	ST	Hawaii: Puu O'umi NAR, Kohalas, 5000ft
waddingtoni[a]	202526a	ST	Maui: East Maui Irrigation, Haiku Uka, Heed Trail, 4200'
waddingtoni[a]	202554b	ST	Maui: Puu Kukui Trail, West Maui
waddingtoni[a]	202415b	ST	Molokai: Puu Kolekole, 3854 ft.
diamphidiopoda[a]	200785a	Anto	Maui
expansa[a]	201012a	BT	Maui
grimshawi[a]	Flybase	PW	Maui

[a] These individuals were also used in the BEST analyses.
[b] For more information on specific collection records contact authors with corresponding barcodes.
[c] Taxonomically defined species group that each specimen has been placed in. ST refers to spoon tarsus subgroup, BT refers to bristle tarsus subgroup, Anto refers to the *antopocerus* subgroup, and PW refers to the picture wing clade.
[d] Island names are in bold.

2.2. DNA amplification and sequencing

All individuals were sequenced for 10 nuclear loci and four mitochondrial loci. Mitochondrial loci were amplified using universal primers (Simon et al., 1994). Primers to amplify nuclear loci were designed by searching Flybase (Tweedie et al., 2009) for genes with exons conserved between *D. grimshawi*, *Drosophila mojavensis* and *Drosophila virilis*, but containing highly divergent introns. Primers spanning these introns were anchored within the conserved exons. We specifically excluded all multicopy genes and members of multigene families. Loci were selected from different chromosome arms or were at least 5 Mbp apart on the same chromosome arm. Gene identity was assessed via BLAST to the annotated *Drosophila melanogaster* and *D. grimshawi* genomes. Details for each locus are listed in Tables 2 and 3.

PCR conditions included an initial denaturing at 95 °C for 5 min, followed by 35 cycles of the following sequence: denaturation at 95 °C for 30 s, annealing between 54 °C and 62 °C (depending on locus) for 30 s, and an extension at 72 °C for 30 s. A final extension step at 72 °C was held for 5 min. PCR products were cleaned using standard ExoSAP-IT (USB) protocols. Cleaned products were sent to the UC Berkeley Sequencing Facility and sequenced in both directions on an ABI 3730 capillary sequencer. Contigs were assembled using Sequencher, ver. 4.7 (GeneCodes, Corp). Because of the recent divergence between the taxa in this study, alignment was trivial and easily performed by eye using MacClade, ver. 4.06 (Maddison

Table 2
Nuclear gene sampling.

Gene[a]	Name	Primer sequence	Number sampled[b]	Linkage group[c]	#Chars (PICs)	Substitution Model[4]	Accession Numbers
PRY	PRYL PRYR	5′-CCGATGTCCTATGGATAGCTTTA-3′ 5′-AAGTGGAACGAATGTGTAAAACG-3′	31	Y	626(24)	HKY	HQ703659–HQ703688
Kl-2	Kl2L kl2R	5′-TAATACAGAACGGTGGTATGGGTAT-3′ 5′-GTTGCTTGGCTAATTCGTAAAGAGT-3′	31	Y	587(14)	HKY + I	HQ703629–HQ703658
Fz4	Fz4L Fz4R	5′-GCGTCTTTCTATTGCGCTACTAT-3′ 5′-GCTTGTACGGACTGCTGATTATT-3′	31	X	1063(39)	SYM + G	HQ703599–HQ703628
Smox	SmxL SmxR	5′-AATTGAAACCGYTSGAGCA-3′ 5′-CTTAGGCATTCGGCAAAGAC-3′	31	X	855(9)	GTR + G	HQ728839–HQ728868
Ge-1	Ge1L Ge1R	5′-ATTGAAATGCAATTGTCCAAACT-3′ 5′-ATACGCATTAAACTCCACCATGA-3′	28	B	767(33)	HKY + I	HQ728723–HQ728749
Pds5	Pds5L Pds5R	5′-GGATACTTTGTGGACAATTCAGAGT-3′ 5′-AGATATTTCACGAACTCTTCAGCAC-3′	31	C	595(28)	HKY + G	HQ728779–HQ728808
Dip3	Dip3L Dip3R	5′-GTTAGAGTGGACATATGGGATCG-3′ 5′-GCACTGTTTCCATACTGTTGTTG-3′	31	C	811(25)	HKY + I	HQ703743–HQ703772
Bin	BinL BinR	5′-CGGCTGGYGCATAATCACTT-3′ 5′-CACAATCTCAGCCTGAACGA-3′	29	D	542(22)	HKY + I + G	HQ728698–HQ728722
Osi9	OsiL OsiR	5′-AGCAGCGGCATCAGRTACTT-3′ 5′-CCCAAGGACTCCATACAGGA-3′	30	E	476(9)	GTR + G	HQ728750–HQ728778
Rad23	RadL RadR	5′-GCGGAATATTTGGTGGAAAA-3′ 5′-GAAGCCACTGTGCGATTGTA-3′	31	F	456(8)	HKY	HQ728809–HQ728838

[a] Gene named after orthologous, annotated gene in *D. melanogaster*.
[b] Number of taxa sampled for this loci. 31 individuals were used in total.
[c] Linkage groups include X and Y sex chromosomes, mitochondrion, or Muller's Elements.
[d] Most likely substitution model for the gene identified by the Akaike Information Criterion (AIC) implemented in MrModeltest v2.3.

Table 3
Mitochondrial gene sampling.

Gene[a]	Name	Primer sequence	Number sampled[b]	Linkage group[c]	#Chars (PICs)	Substitution model[d]	Accession numbers
ND2	192	5′-AGCTATTGGGTTCAGACCCC-3′	31	mt	520(39)	HKY + I + G	HQ703773–HQ703790 HQ170926–HQ170935, HQ170877
	732	5′-GAAGTTTGGTTTAAACCTCC-3′					
COI	2183	5′-CAACATTTATTTTGATTTTTTGG-3′	31	mt	829(89)	GTR + I + G	HQ703707–HQ703724 HQ170816–HQ170826, HQ170766
	3041	5′-TYCATTGCACTAATCTGCCATATTAG-3′					
COII	3037	5′-ATGGCAGATTAGTGCAATGG-3′	31	mt	749(62)	GTR + I + G	HQ703725–HQ703742 HQ170712–HQ170721 HQ170724, HQ170652
	3791	5′-GTTTAAGAGACCAGTACTTG-3′					
16s	16sF	5′-CCGGTTTGAACTCAGATCACGT-3′	31	mt	510(6)	HKY + I	HQ703689–HQ703706 HQ171019–HQ171029, HQ170967
	16sR	5′-CGCCTGTTTAACAAAAACAT-3′					

[a] Gene named after orthologous, annotated gene in *D. melanogaster*.
[b] Number of taxa sampled for this loci. 31 individuals were used in total.
[c] Linkage groups include X and Y sex chromosomes, mitochondrion, or Muller's Element.
[d] Most likely substitution model for the gene identified by the Akaike Information Criterion (AIC) implemented in MrModeltest v2.3.

and Maddison, 2002). All sequences generated in this study have been deposited in GenBank (Table 1).

2.3. Partitioned concatenated analysis

Sequences were concatenated, partitioned and analyzed in Bayesian (mrbayes, ver. 3.1.2; Ronquist and Huelsenbeck, 2003) and maximum likelihood (RAxML, version 7.2.6; Stamatakis, 2006). Maximum likelihood analyses were performed on the Abe Teragrid, accessed through the CIPRES portal (Miller et al., 2009). Likelihood searches were partitioned by locus and the GTRGAMMA model was used for each partition to estimate the tree, since the authors suggest against using proportion of invariant sites and simpler models (Stamatakis, 2006). Two thousand bootstrap replicates were performed to assess support for the inferred relationships. The Akaike information criterion (AIC), implemented in MrModeltest, ver. 2.3 (Nylander, 2004), was used to estimate the best-fit model of substitution for each of the 14 loci in the Bayesian analyses (Tables 2 and 3). The concatenated analysis (with 14 partitions) was run for 10 million generations and was sampled every 100 generations.

2.4. Genealogies

Gene trees derived from individual analyses were estimated using MrBayes. An assumption of the coalescent is that there is recombination between loci. We identified which genes to concatenate into one locus based on their chromosomal locations and amount of recombination between them for the gene tree analyses. Loci on the Y chromosome and mitochondria are known to have no recombination so each was analyzed as one non-recombining unit. Loci residing on the same chromosome arm were tested to ensure recombination was occurring between them using the 4-gamete test (Hudson and Kaplan, 1985) implemented in DnaSP (Librado and Rozas, 2009). All were found to be in linkage equilibrium. Therefore the combined mitochondrial loci, both Y chromosome loci, and every other locus were considered a single non-recombining unit. Each recombining unit was used to estimate a gene tree for a total of 10 gene trees. Gene trees were run for 1 million generations and were sampled every 100 generations. By examining the cumulative split frequencies plot calculated by AWTY (Wilgenbusch et al., 2004) and identifying when the potential scale reduction factor (PSRF) approached 1, convergence in all Bayesian analyses was assessed.

2.5. Bayesian estimation of species trees

BEST, ver. 2.3.1 (Liu et al., 2008) was used to estimate the species tree while taking into account incomplete lineage sorting by minimizing the bifurcations based on deep coalescent events. Allelic data were used to build the gene trees used in this analysis. Alleles for heterozygous sequences were identified using PHASE2.1.1 (Stephens and Scheet, 2005). The alleles identified with a probability greater than 0.5 were kept, and any with lower posterior probabilities were cloned and re-sequenced to confirm the identity of each allele at that gene. BEST assumes gene tree conflict is due to incomplete lineage sorting, and not hybridization. To allow the analysis to reach convergence in a reasonable amount of time, only one allele was used per species (Table 1). Exceptions were made for *Drosophila conformis* and *Drosophila incognita* since they were found not be monophyletic in the partitioned concatenated analysis, and populations of *D. waddingtoni* found on different islands. Since BEST is sensitive to missing data and we were unable to amplify some taxa for the Ge-1 and Bin genes, we removed these loci from the dataset. The partitioned concatenated analysis was reanalyzed without these genes to test the impact they had on the topology, and the only difference was a reduction in support in a few nodes, with no topology change. Gene partitions in this analysis were identified as non-recombining blocks. Substitution models for the combined loci on the Y chromosome and mitochondria were estimated using MrModeltest v2.3 (Nylander, 2004).

Two Markov Chain Monte Carlo searches were run for 45 million generations and were sampled every 1000 generations. Four chains were used for each run, with a heating factor set using temp = 0.10 to allow for adequate mixing. Convergence was estimated using AWTY and PSRF values (Wilgenbusch et al., 2004). θ was set at 0.0023 ($a = 3$, $b = 0.0047$). θ was calculated as the average of the average pair wise divergence for each nuclear allele for *Drosophila dasycnemia*, *Drosophila neutralis*, and *Drosophila sordidapex* from sequences used in this paper.

2.6. Bayesian untangling of concordance knots

BUCKy, ver. 1.3.2 (Ané et al., 2007) was used to estimate the primary concordance tree and assess agreement between different gene trees. Concordance analyses estimate the overall tree quickly and without assuming that conflict is due to any one issue. BUCKy estimates the overall history of the species tree assuming that the dominant signal from independent gene trees is that of the true evolutionary history. This method allows for uncertainty in gene tree estimation and also estimates a level of support by identifying how much of the genome supports each relationship. Since each nuclear locus did not estimate each species as reciprocally monophyletic, the sampling scheme used for the BEST analysis was also used for the BUCKy analysis to circumvent conflict. Two analyses were performed, one with the a priori level of discordance (α) set at 1, the other with α set at 0.1. This allowed for a range of expected discordance, with the analyses with $\alpha = 1$ predicting that most of the gene trees support the species tree, and an $\alpha = 0.1$ predicting up to a different tree per genealogy. Each analysis was run twice for 10 million generations with four chains.

Bayesian concordance analysis infers concordance factors (CF), useful in determining how much support there is for a given topology. CF values can indicate how well the sampled genome corroborates a certain tree (sample CF). Alternatively, genome wide CF values assume that the sampled loci are representative of the genome as a whole, given an assumed level of expected discordance (α). Under these assumptions CF values indicate the degree to which the entire genome is expected to produce a certain topology.

3. Results

3.1. Partitioned analyses

Partitioned, concatenated Bayesian and maximum likelihood analyses produce the same topology, with high levels of support (Fig. 2). The first 2000 trees were discarded as burn in for the partitioned concatenated analysis run in MrBayes based on the differences in split posteriors calculated via AWTY and observed in Tracer. In contrast to the partitioned concatenated analyses, individual gene trees displayed a high degree of conflict and poor resolution, possibly due to the fewer characters in each analysis or conflict due to incomplete lineage sorting or hybridization. Analyses for individual gene trees converged quickly, and the first 1000 trees were discarded as burn in based off of convergence statistics. These figures are available as Supplementary material online.

The partitioned concatenated analysis suggests that the spoon tarsus subgroup as currently defined (Lapoint et al., 2009) is not monophyletic. Two species considered basal within the spoon tarsus subgroup *Drosophila atroscutellata* and *Drosophila fastigata*, are actually basal to the spoon tarsus subgroup plus bristle tarsus

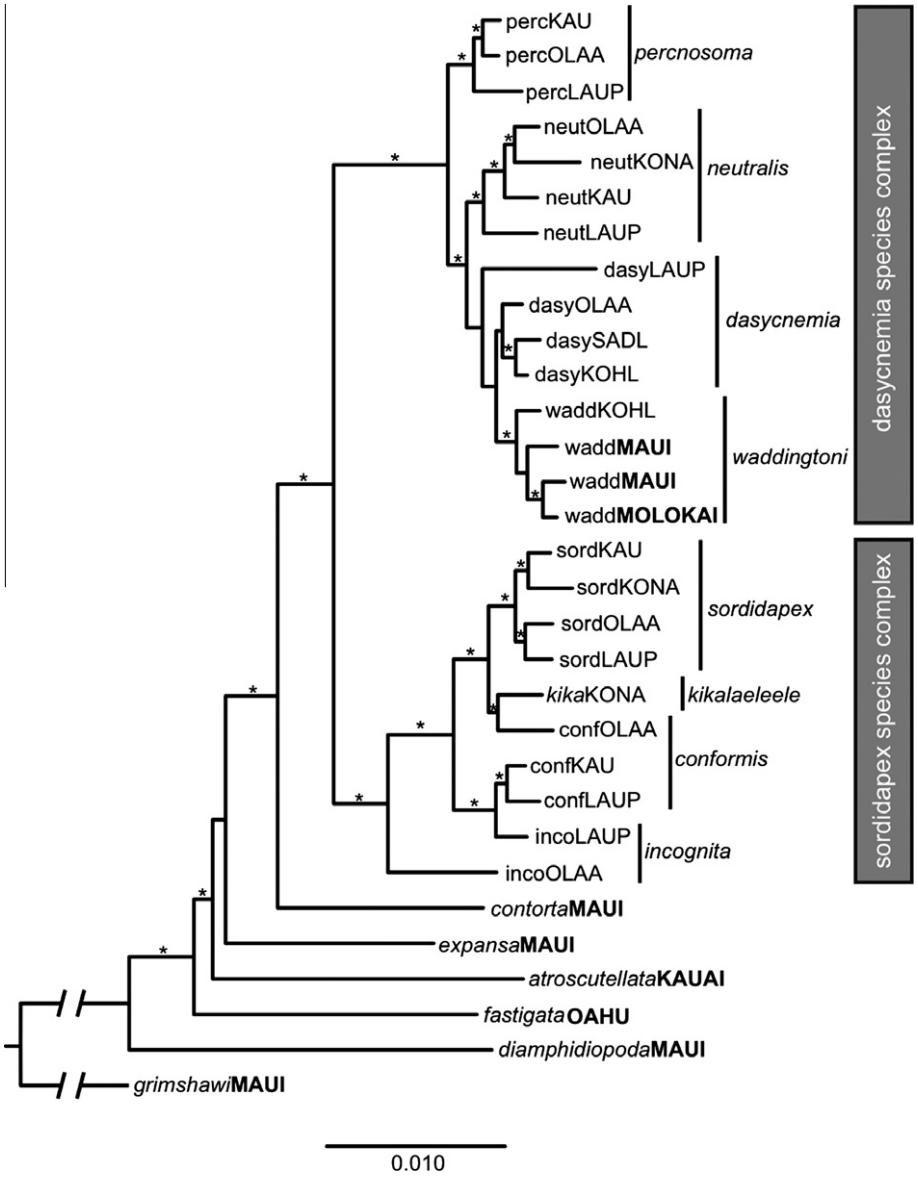

Fig. 2. Bayesian concatenated, partitioned phylogeny of the spoon tarsus subgroup. Maximum likelihood topology is identical the Bayesian topology. Nodes with posterior probabilities greater than 0.90 and bootstrap supports greater than 70% are indicated by an * above branches. All taxa were collected on the Island of Hawaii except where indicated in bold. KAU = Kau Forest Reserve; OLAA = Olaa Tract, HVNP; LAUP = Laupahoehoe NAR; KONA = Kona Forest Reserve; SADL = Upper Waiakea Forest Reserve; KOHL = Puu O'umi NAR.

subgroup, represented by *D. expansa*. This result is in agreement with a larger analysis of all Hawaiian *Drosophila* lineages (O'Grady et al., in press).

The species present on the Big Island form a clade, with the Maui Nui populations of *D. waddingtoni* nested within this larger monophyletic group (Fig. 2). There appear to be two well-supported lineages within Big Island, the *dasycnemia* species complex (sensu Hardy, 1965) that includes *D. dasycnemia*, *D. waddingtoni*, *D. neutralis* and *Drosophila percnosoma* and the *sordidapex* species complex (sensu Hardy, 1965), which includes *D. sordidapex*, *Drosophila kikalaeleele*, *D. conformis*, and *D. incognita* (Fig. 2). While these complexes are supported as monophyletic, several of the species included in each are not. For example, *Drosophila waddingtoni* is nested within a paraphyletic *D. dasycnemia*, although support for the relationships within *D. dasycnemia* are not well supported. Furthermore, both *D. incognita* and *D. conformis* are paraphyletic (Fig. 2).

3.2. Best

The BEST analysis produced a similar phylogeny to the partitioned concatenated analyses (Fig. 3a). The monophyly of the Big Island clade and both the *dasycnemia* and *sordidapex* species complexes are well supported. However, posterior probabilities at the species nodes were reduced in comparison to the partitioned analyses. Species relationships within the *dasycnemia* complex were poorly supported and showed a different order of divergence. Within the *sordidapex* complex species relationships were poorly supported and appear to be recently diverged. The BEST analysis produced a phylogeny that included *D. atroscutellata* and *D. fastigata* within the spoon tarsus subgroup. However, the relationships between *D. atroscutellata*, *D. fastigata* and *D. expansa* were poorly supported (posterior probabilities <0.7) and with very short internodes.

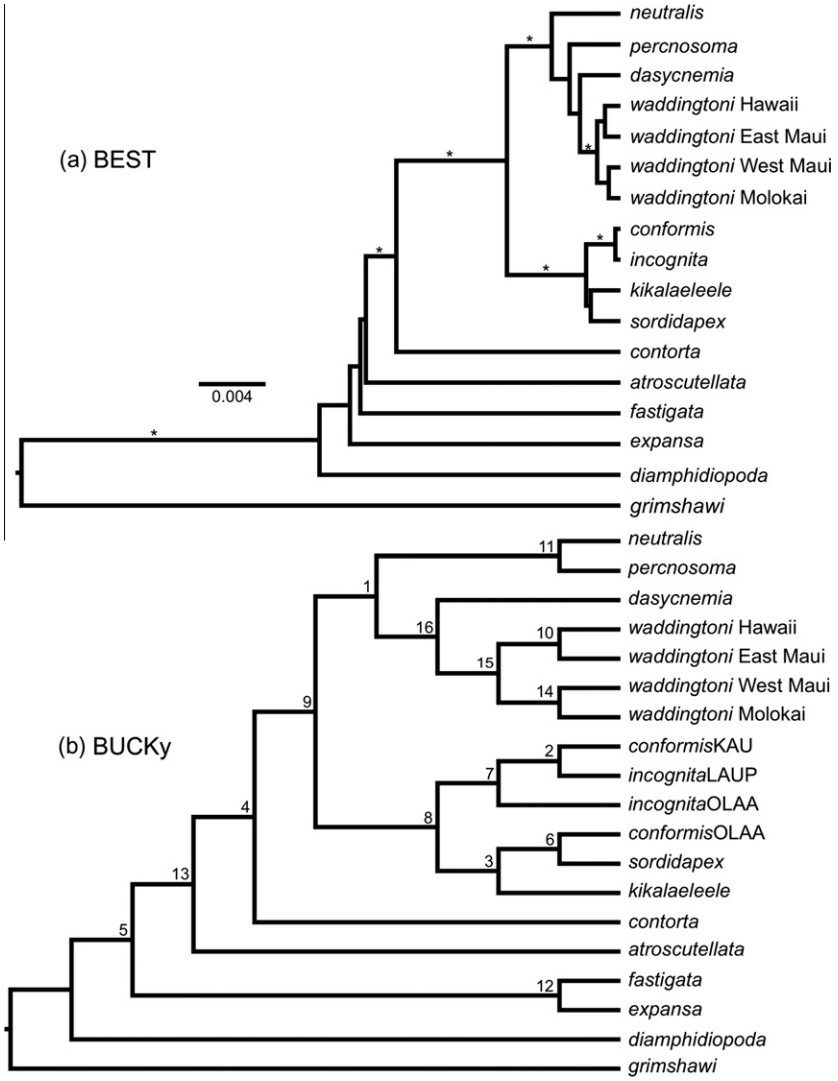

Fig. 3. (a) BEST phylogeny of the spoon tarsus subgroup. Nodes with posterior probabilities greater than 0.90 are indicated by an *. (b) Bayesian concordance analysis of nine loci. Values above branches are primary concordance values, values below are genome wide mean concordance factors for $\alpha = 0.1$ and $\alpha = 1$.

3.3. BUCKy

The Bayesian Concordance analyses estimated relationships within the spoon tarsus subgroup (Fig. 3b) to be very similar to those estimated using concatenation or BEST (e.g., Figs. 2 and 3a). The taxa that were identified as paraphyletic in the concatenated analysis are also found to be paraphyletic in this analysis. The primary concordance values are low throughout the tree, indicating a high level of discord between gene genealogies (see Table 4).

Table 4
BUCKy primary concordance factors.

Splits	Sample-wide[a]	Genome-wide[b] $\alpha = 0.1$	Genome-wide $\alpha = 1.0$
1	0.799 (0.625, 0.875)	0.789 (0.414, 0.990)	0.710 (0.353, 0.955)
2	0.545 (0.375, 0.750)	0.539 (0.189, 0.871)	0.488 (0.167, 0.816)
3	0.409 (0.250, 0.625)	0.404 (0.070, 0.812)	0.364 (0.062, 0.751)
4	0.349 (0.125, 0.500)	0.344 (0.052, 0.717)	0.310 (0.046, 0.662)
5	0.320 (0.125, 0.500)	0.317 (0.038, 0.712)	0.288 (0.035, 0.659)
6	0.313 (0.125, 0.625)	0.310 (0.008, 0.737)	0.282 (0.009, 0.682)
7	0.306 (0.125, 0.500)	0.303 (0.034, 0.677)	0.273 (0.030, 0.624)
8	0.284 (0.125, 0.500)	0.281 (0.008, 0.667)	0.252 (0.007, 0.614)
9	0.279 (0.125, 0.375)	0.276 (0.033, 0.642)	0.248 (0.029, 0.590)
10	0.267 (0.125, 0.500)	0.264 (0.006, 0.671)	0.241 (0.006, 0.619)
11	0.210 (0.000, 0.500)	0.208 (0.000, 0.644)	0.190 (0.000, 0.593)
12	0.202 (0.000, 0.375)	0.200 (0.000, 0.602)	0.183 (0.000, 0.555)
13	0.199 (0.000, 0.375)	0.197 (0.000, 0.580)	0.177 (0.000, 0.532)
14	0.198 (0.125, 0.375)	0.196 (0.007, 0.583)	0.179 (0.007, 0.537)
15	0.192 (0.000, 0.375)	0.189 (0.000, 0.579)	0.171 (0.000, 0.531)
16	0.164 (0.125, 0.250)	0.162 (0.006, 0.511)	0.146 (0.005, 0.469)

[a] Sample wide mean concordance factors with 95% confidence intervals.
[b] Genome wide mean concordance factors with 95% confidence intervals at $\alpha = 0.1$ and $\alpha = 1.0$.

4. Discussion

4.1. Relevance to phylogeny

The placement of the bristle tarsus species, *D. expansa*, in all of these analyses renders the monophyly of the spoon tarsus subgroup, as currently described, suspect. The BUCKy analysis found *D. fastigata* sister to *D. expansa*, making its inclusion in the spoon tarsus subgroup questionable. While the BEST analysis still reconstructs the spoon tarsus subgroup as monophyletic, the internal nodes are very short and poorly supported. This lack of resolution is not entirely unexpected. Whereas all other spoon tarsus species have a second tarsomere characterized by being cuplike filled densely with hair, *D. atroscutellata* and *D. fastigata* do not (Stark

and O'Grady, 2009). These two leaf breeder species have been included in the spoon tarsus subgroup due to the second tarsal segment being merely slightly enlarged and moderately concave. This enlargement and concavity is not nearly as great as the other spoon tarsus species. Indeed, both *D. atroscutellata* and *D. fastigata* both have slight lobes at the apical end of their basitarsi, with some setae. This could cause them to be better included in the bristle tarsus group, though again, these characters are not as developed as other members of that group (Stark and O'Grady, 2009). A revision of the entire AMC – including the *antopocerus* and the modified tarsus species groups – is warranted to further resolve the relationships within this species clade. The rest of the spoon tarsus subgroup is well supported as a monophyletic group in all analyses, including *D. contorta* and the Big Island species, which includes the multi-island *D. waddingtoni*. Unlike *D. atroscutellata* and *D. fastigata*, these species share many common characteristics, including the overall morphology of their spoon and genitalic characters. We feel that the available evidence requires the reexamination of the validity of the inclusion of *D. atroscutellata* and *D. fastigata* in the spoon tarsus subgroup.

Relationships between species vary from well supported and found in all analyses to poorly supported and conflicting. *Drosophila contorta* is well supported as being basal to the rest of the Big Island spoon tarsus species group in all analyses. This is expected based on morphology (Lapoint et al., 2009) and the progression rule (Wagner and Funk, 1995), which predicts basal lineages to be found on older islands. A well-supported split appears to have occurred on the island of Hawaii between the *dasycnemia* and *sordidapex* species complexes. This is also expected given the morphology of these eight species. Members of the *dasycnemia* species complex are generally larger, darker and possess more robust setae on their basitarsi. Conversely, the *sordidapex* species complex is comprised of mostly smaller, light colored flies with patterning on their wings.

Within these species complexes relationships were generally plastic. The BEST analyses found *D. neutralis* as basal to *D. percnosoma*, which was basal to a clade comprised of *D. dasycnemia* and *D. waddingtoni*. In the BEST analysis *D. sordidapex* and *D. kikalaeleele* are sister to each other, though poorly supported. The partitioned concatenated phylogeny estimates *D. percnosoma* as basal to *D. neutralis*, which was basal to *D. dasycnemia* and *D. waddingtoni*. *Drosophila conformis* and *D. incognita* render the *sordidapex* complex paraphyletic in the BUCKy and partitioned concatenated analysis, and the BEST analysis finds the species within this complex poorly supported and recently diverged.

The monophyly of the species found on the Big Island was tested using multiple individuals per species from multiple populations using partitioned concatenated analyses and paraphyletic taxon were included in both the BUCKy and BEST analyses. The species in the *dasycnemia* complex are well-supported monophyletic groupings, indicating that there has been enough time since divergence for lineage sorting and subsequently high species identity. *Drosophila percnosoma*, *D. neutralis* and *D. waddingtoni* were all found to be reciprocally monophyletic. Despite being found on multiple islands, *D. waddingtoni* was shown to be a recently derived species nested within *D. dasycnemia*.

The *sordidapex* species complex is much more complicated, exhibiting high levels of gene tree/species tree conflict. *Drosophila sordidapex* is monophyletic, but *D. conformis* and *D. incognita* are found throughout this species group. The rarity of *D. kikalaeleele* prevented the collection of more than one individual, and could not be tested. Despite being morphologically diagnosable, the polyphyly in the *sordidapex* species complex indicates possibly porous barriers to gene flow, recent diversification, and an overall reduced differentiation. This group may represent lineages that have not yet begun independent evolutionary trajectories, but increased gene and taxon sampling should improve estimation of gene flow and species limits.

4.2. Biogeographic implications

Hawaiian *Drosophila* diversification has been characterized as being driven by mating behavior and ecological adaption in addition to geographic isolation, but the most obvious pattern has been seen in the progression rule. Older lineages are generally found on older islands and younger lineages on younger islands. This pattern appears to be recapitulated here since *D. fastigata* and *D. atroscutellata* are found on the oldest islands, Oahu and Kauai respectively, *D. contorta* and *D. expansa* are from Maui, and the youngest lineages are found on the youngest island, Hawaii.

In addition to the progression down the island chain we find evidence for recent back colonization up the island chain. *Drosophila waddingtoni* is unique for being a multi-island endemic (Nitta and O'Grady, 2008) and for originating on the Island of Hawaii and back colonizing the islands of Maui Nui, in apparently a stepping stone pattern from the youngest volcano on Eastern Maui, then up to the older volcanoes of Maui and Molokai. This pattern is the opposite of the progression rule prevalent in the Hawaiian *Drosophila* (Bonacum, 2001; Bonacum et al., 2005; O'Grady et al., in press) and is not expected since they would have to invade very complex and mature ecosystems (Gillespie et al., 2008). Further exploration of the biogeography of this species is warranted to identify the nature of this interesting back colonization.

4.3. Comparison of methods

Large data sets containing many independently evolving loci are becoming more common in phylogenetic inference, especially in recent radiations like Hawaiian *Drosophila*. However, using multiple loci means that the very real issue of genealogical conflict must be addressed. A number of computational methods have recently become available to incorporate incomplete lineage sorting into phylogenetic reconstruction. These methods are improving our understanding of evolutionary relationships within groups that have recently and rapidly diverged and for which morphological or single gene phylogenies were unable to resolve relationships. These are allowing researchers to address important, yet computationally difficult, evolutionary questions within robust phylogenetic context.

This analysis improves on the most recent estimation of this group that uses solely mitochondrial sequences to infer the phylogeny of the entire Hawaiian *Drosophila* (O'Grady et al., in press). While mitochondrial sequences are useful in phylogenetics due to the ease of sequence generation via universal primers (Simon et al., 1994), high variability (Moritz et al., 1987) and rapid lineage sorting as a result of their maternal mode of inheritance (Avise, 2004), the non-recombining nature of mtDNA means that all mitochondrial loci present the same evolutionary history, a history that may not reflect actual relationships among species. Furthermore, incomplete lineage sorting, introgression and selection can obscure the phylogenetic signal present in this one marker and can lead to complications when estimating phylogeny (Maddison, 1997; Leache and McGuire, 2006). The addition of the nuclear loci greatly improves the estimation of the species relationships within the spoon tarsus subgroup since they represent multiple independent genealogies (Edwards, 2009). This should improve the inference of the phylogeny under the assumptions of total evidence (Kluge, 1989), the true signal should swamp out misleading signal caused by demography and selection (Rokas et al., 2003). In addition, by exploring analyses like BEST that use the amount of conflict between genealogies as a source of information, multiple loci should improve phylogenetic estimation (Knowles, 2009).

The partitioned concatenated analyses and BEST analysis differed in several important aspects. First, the concatenated analyses displayed higher resolution and higher levels of support, possibly due to the larger numbers of informative sites. BEST analyzes the species tree as a sum of the individual gene trees, which were poorly resolved due to a lack of informative characters per gene, and this is reflected in the phylogeny. Since one gene is unlikely contain this much information this support and resolution may therefore be artificially inflated in concatenated analyses, particularly if the actual history of the species tree is comprised of very recent and simultaneous divergences that should be poorly supported given the data. BEST infers phylogeny in a more biologically realistic way for young lineages by taking into account the process of incomplete lineage sorting.

The BEST and concordance analyses were similar, but also differ is some respects. BEST assumes species are reciprocally monophyletic, unlikely in very recent radiations. This is also detrimental when one of your goals is testing for species level monophyly as we are. BUCKy is much less computationally intensive, with gene trees being inferred independently from the species tree, while BEST not only simultaneously estimates gene trees and a species tree but many other parameters as well, which can make the analyses prohibitively long. Concordance analyses also do not assume what is the cause of conflict between species, and just attempts to minimize this (Ané et al., 2007). This can be helpful if conflict is expected to be due to more than just retention of ancestral polymorphism, but conversely is less biologically defensible since it is not modeling any method. BEST improves on BUCKy in this way at taxonomic levels that are not confounded by gene flow, and support on these topologies shows this.

4.4. Causes of conflict

There are several reasons species phylogenies are to difficult to directly infer from genealogies: incomplete lineage sorting obscuring species level relationships by drift; introgression of alleles from one lineage to another; and cryptic species can assumptions of monophyly. When multiple loci are analyzed in either a concatenated or coalescent approach, *D. incognita* and *D. conformis* are still found to be paraphyletic or influence the species tree to the point of reducing support for other species relationships. There is a high degree of conflict between the different genealogies in our current analysis, possibly due to any of these issues.

Cryptic species are morphologically indistinguishable species that represent distinct evolutionary lineages (Bickford et al., 2006). The *sordidapex* complex is comprised of nondescript spoon tarsus species with only a few defining characters to discern each and the paraphyly in this group could be attributed to morphologically similar species being described as a single species when in fact they represent distinct lineages. One *D. incognita* specimen is on a relatively distinct branch in the partitioned concatenated analysis, and is the most likely candidate for a cryptic species. The rest of the species in this complex appear to suffer from a shared evolutionary history, instead of representing distinct lineages.

Introgression is a possible cause of the genealogical conflict in the spoon tarsus. Picture wing Hawaiian *Drosophila* species have been shown to be capable of hybridization in wild and laboratory settings (Yang and Wheeler, 1969; Carson et al., 1989), but while *D. dasycnemia* displays to *D. sordidapex* and vice versa, these displays do not lead to copulation (Spieth, 1966). Hybrid phenotypes have not been documented but the partially overlapping mating behaviors of the different species of spoon tarsus (Spieth, 1966) still make hybridization a possibility. The individual genealogies show widespread admixture in both the *sordidapex* and *dasycnemia* complexes, but this could be due to either incomplete lineage sorting or introgression. Differentiating introgression and incomplete lineage sorting can be difficult in recently divergent species since both produce the same pattern of shared polymorphisms between morphologically identifiable species (Holder et al., 2001). A more exhaustive study on the permeability of species boundaries in the spoon tarsus subgroup is warranted.

Despite the possibility of cryptic species and introgression we believe that incomplete lineage sorting in a very young radiation with large population sizes is the most likely cause of the conflict. We assume the *dasycnemia* and *sordidapex* complexes of the spoon tarsus subgroup are no more than half a million years old, since the Island of Hawaii could not have been colonized before then (Price and Clague, 2002). In such recent radiations it is possible that drift will cause more genealogies to misrepresent the phylogenetic history of the species than those that corroborate it (Knowles and Carstens, 2007). The methods employed in this study address this issue of incomplete lineage sorting and find that there has not been enough time for lineages to coalesce into discrete lineages in the *sordidapex* complex.

Acknowledgments

I would like to thank the Hawaii Natural Area Reserves and Department of Fish and Wildlife for their help with permitting. Thanks to K. Magnacca, G. Bennett, B. Ort, D. Crowser, S. Bridgers, K. Tran, E. Young, N. Pantoja, E. Owen, and J. Eldon for help with specimen collections and advice. Thank you to two anonymous reviewers whose comments have greatly improved this manuscript. This work was supported by NSF DEB 0842348 and the UC, Berkeley Walker Fund.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2010.12.015.

References

Ané, C., Larget, B., Baum, D.A., Smith, S.D., Rokas, A., 2007. Bayesian estimation of concordance among gene trees. Mol. Biol. Evol. 24, 412–426.
Avise, J.C., 2004. Molecular Markers, Natural History, and Evolution, second ed. Sinauer Associates, New York.
Baker, R.H., DeSalle, R., 1997. Multiple sources of character information and the phylogeny of Hawaiian drosophilids. Syst. Biol. 46, 654–673.
Baldwin, B., Sanderson, M.J., 1998. Age and rate of diversification of the Hawaiian silversword alliance. Proc. Natl. Acad. Sci. USA 95, 9402–9406.
Belfiore, N.M., Liu, L., Moritz, C., 2008. Multilocus phylogenetics of a rapid radiation in the genus Thomomys. Syst. Biol. 57, 294–310.
Bickford, D., Lohman, D.J., Sodhi, N.S.P.K.L., Ng Meier, R., Winke, K., Ingram, K.K., Das, I., 2006. Cryptic species as a window on diversity and conservation. Trends Ecol. Evol. 22, 148–155.
Bonacum, J., 2001. Molecular Systematics of the Hawaiian Drosophilidae. Phd Thesis, Yale University, New Haven, Connecticut.
Bonacum, J., O'Grady, P.M., Kambysellis, M., DeSalle, R., 2005. Phylogeny and age of diversification of the *planitibia* species group of the Hawaiian *Drosophila*. Mol. Phylogenet. Evol. 37, 73–82.
Brandley, M.C., Schmitz, A., Reeder, T.W., 2005. Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. Syst. Biol. 54, 373–390.
Brumfield, R.T., Liu, L., Lum, D.E., Edwards, S.V., 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.
Carson, H.L., Clague, D.A., 1995. Geology and biogeography of the Hawaiian Islands. In: Wagner, W., Funk, V. (Eds.), Hawaiian Biogeography: Evolution in a Hotspot Archipelago. Smithsonian Institution Press, Washington, DC, pp. 14–29.
Carson, H.L., Stalker, H.D., 1969. Polytene chromosome relationships in Hawaiian species of *Drosophila* IV. The *D. primaeva* subgroup. Univ. Tex. Publ. 6918, 85–94.
Carson, H.L., Kaneshiro, K.Y., Val, F.C., 1989. Natural hybridization between the sympatric Hawaiian species *Drosophila silvestris* and *Drosophila heteroneura*. Evolution 43, 190–203.
Degnan, J.H., Rosenberg, N.A., 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2, 762–768.

Drosophila 12 Genomes Consortium, 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. Nature 450, 203–218.
Edwards, S.V., 2008. A smörgåsbord of markers for avian ecology and evolution. Mol. Ecol. 17, 945–946.
Edwards, S.V., 2009. Is a new and general theory of molecular systematics emerging? Evolution 63, 1–19.
Eldredge, L.G., Evenhuis, N.L., 2003. Hawai'i's biodiversity: a detailed assessment of the numbers of species in the Hawaiian Islands. Bishop Museum Occasional Papers 76, 1–30.
Fleischer, R.C., McIntosh, C.E., Tarr, C.L., 1998. Evolution on a volcanic conveyor belt: using phylogeographic reconstructions and K–Ar based ages of the Hawaiian Islands to estimate molecular evolutionary rates. Mol. Ecol. 7, 533–545.
Freeland, J.R., Boag, P.T., 1999. The mitochondrial and nuclear genetic homogeneity of the phenotypically diverse Darwin's ground finches. Evolution 53, 1553–1563.
Gadagkar, S.R., Rosenberg, M.S., Kumar, S., 2005. Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. J. Exp. Zool. 304B, 64–74.
Gillespie, R.G., Claridge, E.M., Roderick, G.K., 2008. Biodiversity dynamics in isolated island communities: interaction between natural and human-mediated processes. Mol. Ecol. 17, 45–57.
Givnish, T.J., Sytsma, K.J., Patterson, T.A., Hapeman, J.R., 1996. Comparison of patterns of geographic speciation and adaptive radiation in Cyanea and Clermontia (Campanulaceae) based on a cladistic analysis of DNA sequence and restriction-site data. Am. J. Bot. 83, 159.
Hardy, D.E. 1965. Insects of Hawaii. Diptera: *Cyclorrhapha* II, Series *Schizophora*, Section Acalypterae I. Family Drosophilidae, vol. 12. University of Hawaii Press, Honolulu, 814 pp.
Heed, W.B., 1968. Ecology of the Hawaiian drosophilidae. Univ. Tex. Publ. 6818, 388–419.
Holder, M.T., Anderson, J.A., Holloway, A.K., 2001. Difficulties in detecting hybridization. Syst. Biol. 50, 978–982.
Hudson, R.R., Kaplan, N., 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111, 147–164.
Jordan, S., Simon, C., Polhemus, D., 2003. Molecular systematics and adaptive radiation of Hawaii's endemic damselfly genus megalagrion. Syst. Biol. 52, 89–109.
Kambysellis, M.P., Ho, K.F., Craddock, E.M., Piano, F., Parisi, M., Cohen, J., 1995. Pattern of ecological shifts in the diversification of Hawaiian *Drosophila* inferred from a molecular phylogeny. Curr. Biol. 5, 1129–1139.
Kluge, A.G., 1989. A concern for evidence and a phylogenetic hypothesis of relationships among Epicrates (Boidae, Serpentes). Syst. Zool. 38, 7–25.
Knowles, L.L., 2009. Estimating species trees: methods of phylogenetic analysis when there is incongruence across genes. Syst. Biol. 58, 463–467.
Knowles, L.L., Carstens, B.C., 2007. Delimiting species without monophyletic gene trees. Syst. Biol. 56, 400–411.
Kubatko, L.S., 2009. Identifying hybridization events in the presence of coalescence via model selection. Syst. Biol. 58 (5), 478–488.
Lapoint, R.T., Magnacca, K.N., O'Grady, P.M., 2009. Review of the *spoon tarsus* subgroup of Hawaiian *Drosophila* (Diptera: Drosophilidae), with a description of one new species. Zootaxa 2003, 53–68.
Leache, A.D., McGuire, J.A., 2006. Phylogenetic relationships of horned lizards (*Phrynosoma*) based on nuclear and mitochondrial data: evidence for a misleading mitochondrial gene tree. Mol. Phylogenet. Evol. 39, 628–644.
Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452.
Liu, L., Pearl, D.K., Brumfield, R.T., Edwards, S.V., 2008. Estimating species trees using multiple-allele DNA sequence data. Evolution 62, 2080–2091.
Lopez-Fernandez, H., Winemiller, K.O., Honeycutt, R.L., 2010. Multilocus phylogeny and rapid radiations in neotropical cichlid fishes (Perciformes: Cichlidae: Cichlinae). Mol. Phylogenet. Evol. 55 (3), 1070–1086.
Maddison, W.P., 1997. Gene trees in species trees. Syst. Biol. 46, 523–536.
Maddison, D., Maddison, W., 2002. MacClade: Analysis of Phylogeny and Character Evolution. Sinauer Associates, Sunderland, Massachusetts.
Magnacca, K.N., Foote, D., O'Grady, P.M., 2008. A review of the endemic Hawaiian drosophilidae and their host plants. Zootaxa 1728, 1–58.
McCormack, J.E., Huang, H., Knowles, L.L., 2009. Maximum-likelihood estimates of species trees: how accuracy of phylogenetic inference depends upon the divergence history and sampling design. Syst. Biol. 58 (5), 501–508.
Miller, M.A., Holder, M.T., Vos, R., Midford, P.E., Liebowitz, T., Chan, L., Hoover, P., Warnow, T., 2009. The CIPRES Portals, CIPRES.
Montgomery, S.L., 1975. Comparative breeding site ecology and the adaptive radiation of picture-winged *Drosophila* (Diptera: Drosophilidae) in Hawaii. Proc. Hawaiian Entomol. Soc. 22, 65–103.
Moritz, C., Dowling, T.E., Brown, W.M., 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annu. Rev. Ecol. Syst. 18, 269–292.
Nitta, J.H., O'Grady, P.M., 2008. Mitochondrial phylogeny of the endemic Hawaiian craneflies (Diptera, Limoniidae, *Dicranomyia*): implications for biogeography and species formation. Mol. Phylogenet. Evol. 46 (3), 1182–1190.
Nylander, J.A.A., 2004. MrModeltest v2. Program Distributed by the Author. Evolutionary Biology Centre, Uppsala University.
O'Grady, P.M., Zilversmit, M., 2004. Phylogenetic relationships within the *haleakalae* species group inferred by molecular and morphological characters (diptera: drosophilidae). Bishop Museum Bull. Entomol. 12, 117–134.
O'Grady, P.M., Magnacca, K.N., Lapoint, R.T., 2010. Taxonomic relationships within the endemic Hawaiian Drosophilidae. Rec. Hawaii Biol. Surv. 108, 3–35.
O'Grady, P.M., Lapoint, R.T., Bonacum, J., Lasola, J., Owen, E., Wu, Y., DeSalle, R., in press. Phylogenetic and ecological relationships of the Hawaiian *Drosophila* inferred by mitochondrial DNA analysis. Mol. Phylogenet. Evol.
Pollard, D.A., Iyer, V.N., Moses, A.M., Eisen, M.B., 2006. Widespread discordance of gene trees with species tree in Drosophila: evidence for incomplete lineage sorting. PLoS Genet. 2, e173.
Price, J.P., Clague, D.A., 2002. How old is the Hawaiian biota. Geology and phylogeny suggest recent divergence. Proc. Roy. Soc. Lond. B 269, 2429–2435.
Rokas, A., Williams, B.L., King, N., Carroll, S.B., 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425, 798–804.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes. 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
Russo, C.M.A., Takezaki, N., Nei, M., 1995. Molecular phylogeny and divergence times of Drosophilid species. Mol. Biol. Evol. 12, 391–404.
Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Am. 87, 651–701.
Spieth, H.T., 1966. Courtship behavior of endemic Hawaiian *Drosophila*. Univ. Tex. Publ. 6615, 245–313.
Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.
Stark, J.B., O'Grady, P.M., 2009. Morphological variation in the forelegs of Hawaiian Drosophilidae. I The AMC clade. J. Morphol. 271, 86–103.
Stephens, M., Scheet, P., 2005. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. Am. J. Human Genet. 76, 449–462.
Thomson, R.C., Shedlock, A.M., Edwards, S.V., Shaffer, H.B., 2008. Developing markers for multilocus phylogenetics in non-model organisms: a test case with turtles. Mol. Phylogenet. Evol. 49 (2), 514–525.
Tweedie, S., Ashburner, M., Falls, K., Leyland, P., McQuilton, P., Marygold, S., Millburn, G., Osumi-Sutherland, D., Schroeder, A., Seal, R., et al., 2009. Flybase: enhancing *Drosophila* gene ontology annotations. Nucleic Acids Res. 36, D555–D559.
Wagner, W.L., Funk, V.A., 1995. Hawaiian Biogeography, Evolution on a Hotspot Archipelago. Smithsonian Institution, Washington.
Wahlberg, N., Wheat, C.W., 2008. Genomic outposts serve the phylogenomic pioneers: designing novel nuclear markers for genomic DNA extractions of lepidoptera. Syst. Biol. 57, 231–242.
Wilgenbusch, J.C., Warren, D.L., Swofford, D.L., 2004. AWTY: A system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference.
Wirth, W.W., 1952. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. Proc. Haw. Entomol. Soc. 14, 443–484.
Yang, H.Y., Wheeler, M.R., 1969. Studies on interspecific hybridization within the picture-winged group of endemic Hawaiian *Drosophila*. Univ. Tex. Publ. 6918, 131–170.