Molecular Phylogenetics and Evolution 109 (2017) 388–403

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Sex-biased secondary contact obscures ancient speciation onto relictual host trees in central California moths (*Syndemis*: Tortricidae)

Daniel Rubinoff [a,*], Michael San Jose [a], Jerry A. Powell [b]

[a] University of Hawai'i at Manoa, Department of Plant and Environmental Protection Sciences, 3050 Maile Way, Gilmore 310, Honolulu, HI 96822, United States
[b] University of California, Berkeley, Essig Museum of Entomology, 214 Wellman Hall, Berkeley, CA 94720-3112, United States

ARTICLE INFO

Article history:
Received 19 October 2016
Revised 18 January 2017
Accepted 20 January 2017
Available online 22 January 2017

Keywords:
Syndemis
mtDNA
Introgression
Incongruent genomes
Sequoia
Pinus radiata
Lepidoptera

ABSTRACT

The tortricid moth genus *Syndemis* has ten described species, with two polyphagous species in Europe and North America respectively. We sequenced five nuclear and four mitochondrial genes for *Syndemis* samples across both continents and discovered unexpected, extensive diversification restricted to California. DNA evidence supports five new, undescribed, species endemic to California, while the rest of North America and Europe have only one species each. Further, the nuclear genes are less variable and yield contrasting phylogenetic signal compared to mitochondrial DNA for basal relationships between taxa across the genus. Such conflict strongly suggests that male and female moths exhibit radically different levels of philopatry. Our results highlight the importance of sex-specific behavior, and the need for inclusion of multiple genes to fully understand species boundaries, their causes, and the process of speciation. While mtDNA introgression often is invoked to explain incongruous haplotype distributions, our study shows that nuclear DNA selective sweeps, or swamping, can occur while mtDNA and ecology preserve an ancient divergence that is not discernable in nuclear DNA. This study further demonstrates that diversification of herbivores may occur on relictual, declining hostplants, which contrasts with the dominant co-speciation model.

1. Introduction

Herbivorous insect diversity is thought to be very high because of co-diversification with their plant hosts (e.g. Farrell, 1998), but the mechanisms for this phenomenon and its co-evolutionary underpinnings are still the topic of active debate (Janz, 2011); host plant specialization and reproductive isolation must be important in the incipient stages, though detailed evidence for this is lacking. Any evidence informing patterns of diversification in closely-related herbivorous insects should help to reveal the functional basis for this global principle. In fact, herbivorous groups that do not follow this pattern may be more useful as contrasts for understanding how explosive co-diversification might function in ultra-diverse groups.

Because mtDNA usually is maternally inherited and undergoes limited recombination, it can produce different phylogenetic relationships from those reflected in nDNA (Leaché, 2010; Roe and Sperling, 2007; Rubinoff et al., 2006; Shaw, 2002). Incomplete lineage sorting (ILS) between mtDNA and nDNA frequently is the basis for such incongruity. Over time, gene lineages may sort between taxa (Funk and Omland, 2003) suggesting that ILS is a symptom of more recent divergence. However, if the biology of a species supports independence of the two genomes through sex-biased dispersal and divergent ecological pressure, mtDNA and nDNA may have independent evolutionary trajectories over longer periods of time and come to reflect perpetually different evolutionary histories within the same individual, or population.

Defining species and speciation can be ambiguous; conflicting information from different genomes confounds clear boundaries and obscures the definition and recognition of independent lineages (Degnan and Rosenberg, 2009; Funk and Omland, 2003; Rubinoff, 2006; Rubinoff et al., 2006). Specifically, a rigid theoretical dogma of 'species' often is compromised by phenomena such as incomplete sorting of genes between sibling taxa, retained ancestral polymorphisms and hybridization. ILS may result in conflict between genomes, which can be robust and reflect clear, but different, evolutionary histories. In essence, this reveals an arbitrary nature to the recognition of species in such cases, unless additional data are available to elucidate the nature and relevance of the genomic conflict. Choosing to recognize or ignore genetic incongruity in the course of recognizing species should be informed by additional sources of data, e.g. morphology and ecology (Will

* Corresponding author.
E-mail addresses: rubinoff@hawaii.edu (D. Rubinoff), mdsjose@hawaii.edu (M. San Jose), powellj@berkeley.edu (J.A. Powell).

http://dx.doi.org/10.1016/j.ympev.2017.01.010
1055-7903/

et al., 2005; Rubinoff, 2006) so that the species concept employed is as synergistic as possible.

The holarctic tortricid moth genus *Syndemis* includes ten species (Brown, 2005); one widespread species is described from the Nearctic and another from the Palearctic, the remainder have restricted ranges in the Southeast Asian and Oriental Regions. In eastern North America, larvae of *S. afflictana* (Walker) are recorded feeding on diverse trees and shrubs, including conifers and angiosperms (Prentice, 1966; Chapman and Lienk, 1971). In California, however, *S. afflictana* appears to be restricted to coniferous hosts. In the Palearctic, *S. musculana* (Hübner) is widespread from Europe (Scandinavia to the Mediterranean) through Siberia and Japan, in boreal habitats. It is also polyphagous, feeding on a wide variety of angiosperm and coniferous host plants (Bradley et al., 1973).

During the past 40 years, isolated populations of *Syndemis* were discovered along the coast of Central California by several lepidopterists. Some of these are closely associated with relict stands of redwood (*Sequoia sempervirens*), which have persisted in coastal fog belt microclimates, possibly since the mid-to late-Tertiary, when the redwoods became isolated by reduced rainfall and increasing summer temperatures inland (Axelrod, 1976). During cooler, wetter periods, redwoods fossil evidence suggests that the trees were widely distributed across much of western North America. The adult moths of the redwood-feeding *Syndemis* are active during the day primarily from late May-July, which is very unusual for the genus, since *S. afflictana* and *S. musculana* fly once, in very early spring, and both are nocturnal.

Other *Syndemis* populations in coastal central California occur in small endemic stands of Monterey pine (*Pinus radiata*). This pine species is a commercial timber tree in Portugal and dominant in the timber industry across the Southern Hemisphere (Burgess and Wingfield, 2002), but occurs natively only in three isolated populations along the California coast. Originally evolving from a group of closed-cone pines in Mexico, Monterey pine is thought to have been carried north on the Pacific Plate, now isolated from close relatives and restricted to foggy coastal bluffs and valleys (Axelrod, 1967). Whereas other *Syndemis* populations across Europe and North America feed on a variety of host plants, both of the central California populations are restricted to relictual coniferous hosts. Moreover, every other population of *Syndemis* in North America is single brooded, flying in early spring; as part of this study and through museum specimens, we discovered that adults of the central California population on *P. radiata* are recorded from April-May and August-December. The differences in flight period and lack of host specialization of *Syndemis* across most of North America and Europe as compared to the isolated, host-specific central California species demonstrates an extraordinary variability and flexibility in the evolution of life history traits within a small genus. In many other moth genera, for example, host use and adult activity are much more strictly canalized within groups (eg Rubinoff and Sperling, 2002). Understanding the relationships of these atypical populations which are isolated on relictual, declining, host plants will contribute to our understanding of the importance of ecology in the process of speciation. Specifically, we ask how the central coast redwood and Monterey pine feeding species are related to each other (the two species are interspersed like alternating beads on a string hugging the coast), and to *Syndemis* in the North Coast, Cascade and Sierra ranges of California. One hypothesis is that, like the rest of North America and Europe, they represent one genetically contiguous population, with variations in hostplant use. Another is that California is a global center of *Syndemis* diversity, and each region maintains unique, diagnosable, species, to the extent that the central coast supports geographically paraphyletic populations. Specifically, how are these California populations related to *Syndemis* from across North America and Europe? Genitalia dissections, which usually provide diagnostic evidence of species boundaries in tortricine moths (eg Horak, 1999), have revealed no robust differences among any of the populations of *S. musculana*, *S. afflictana*, or the coastal Californian species.

Because we have collected ecological data on many of the populations and morphological data, including informative wing pattern information, for all the Nearctic species, this study system presents a special opportunity to examine the interplay between genomes, ecological factors such as host plant specialization and flight period and geography in the process of speciation. *Syndemis*, particularly with respect to the isolated central California populations, represents an ideal framework for understanding species boundaries and their corresponding phylogenetics in light of inconsistent ecological and morphological divergence.

2. Materials and methods

2.1. Taxon sampling and morphology

We sampled *Syndemis* moths using light traps and pheromone lures (Ret-C Lure, [Daterman et al., 1995; Powell and De Bendicts, 1995]) from Tennessee, Idaho, Alberta, and extensively in northern California, as well as northern, central and southeastern Europe (Fig. 1). Taxa, localities, and voucher codes are listed in Table 1. Adult specimens from all regions were dissected for genitalic comparisons. Protocol for genitalia dissections follows Brown and Powell (2000) except we used euparal as the mounting medium in recent years. Larvae of *Syndemis* were collected by visual search of lower limb foliage of *Sequoia* and were reared to adulthood in $12 \times 18''$ polyethylene bags lined with paper toweling.

2.2. DNA Extraction, amplification and sequencing

For each specimen, head and thorax were removed and used for total genomic DNA extraction. The remainder of the specimen was deposited as a voucher in the University of Hawaii Insect Museum (UHIM). Genomic DNA was extracted using the DNeasy animal blood and tissue extraction kit following recommended protocols (Qiagen, Inc., Valencia, CA).

Nine different gene regions were amplified for this study (Table 2). Five nuclear genes (28s-rDNA, CAD, EF-1a, MDH and RPS5) and four mitochondrial genes (COI, COII, COIII and NAD1) were chosen because each has been demonstrated to be informative in distinguishing populations, species, or genera in various lineages of Insecta (Hillis and Dixon, 1991; Folmer et al., 1994; Simon et al., 1994, 2006; Cho et al., 1995; Moulton and Wiegmann, 2004; Wahlberg and Wheat, 2008). All genes were amplified under the following conditions: 2 min at 94 °C, 40 cycles of (94 °C for 30 s, annealing temperature optimized for each primer pair for 30 s, and 70 °C for 1 min) with a final 70 °C extension for 10 min. Primer sequences and annealing temperatures are listed in Table 2. PCR-products were visualized on 1% agarose gel and purified using QIAquick® spin columns (Qiagen, Inc., Valencia, CA) according to the manufacturer's protocol. Bidirectional DNA sequencing was performed at the Advanced Studies in Genomics, Proteomics and Bioinformatics (ASGPB) sequencing facility of the University of Hawaii at Manoa (http://asgpb.mhpcc.hawaii.edu/). Sequences were deposited into Genbank under accession numbers KY501245-KY501635.

2.3. Sequence alignment, nucleotide composition and phylogenetic analysis

Sequence contig assemblies were aligned by comparison with reference sequences obtained from Genbank with the software

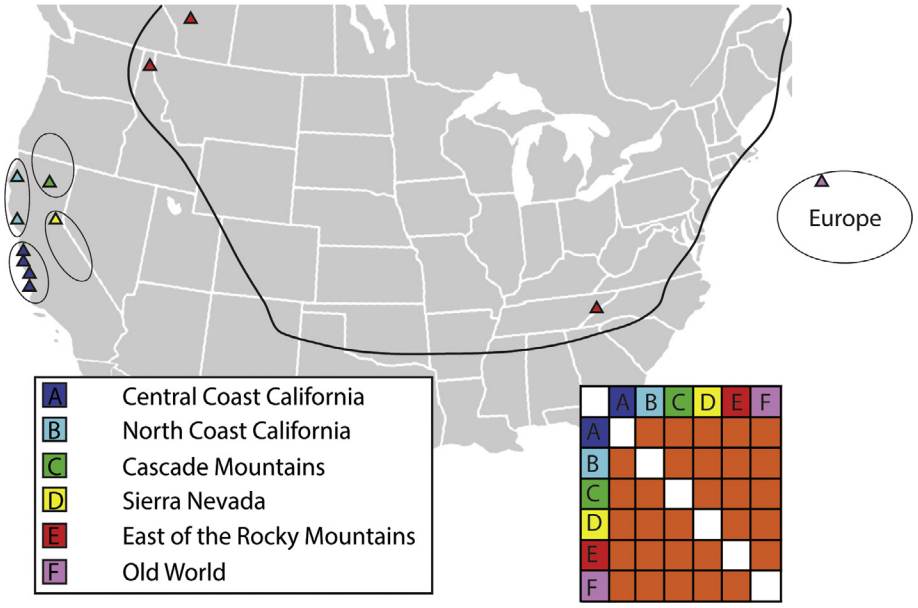

Fig. 1. Map of United States showing species ranges. Color and codes for each range are for biogeographic analyses. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

package Geneious v6.0.5 (Biomatters). Sequence contigs in Geneious were aligned using the "Muscle" option with default settings (Edgar, 2004) to create sequence alignments for each gene and subsequently checked by eye for any misaligned sequences. There was no variation in length between sequences within each locus. We used PAUP∗ 4.0 (Swofford, 2003) to explore rate variation between genes and taxa (uncorrected pairwise difference and number of parsimony informative sites) however due to the complete lack of genetic variation in 28s-rDNA sequences for *Syndemis* across our samples this gene was not included in the analyses. The remaining eight gene regions were tested separately using jModeltest (Darriba et al., 2012) for an appropriate nucleotide substitution pattern under Akaike information criterion with correction (AICc). We also designated the codon position for each gene and tested the data as a whole using PartitionFinder v1.1.1 (Lanfear et al., 2012, 2014). PartitionFinder selects the best-fit partition schemes and models of evolution for phylogenetic analysis using a Bayesian information criterion (BIC). Partition schemes and models of evolution for each partition found using PartitionFinder and the model of evolution for each gene found using jModeltest found are in Tables 3and 4. Phylogenetic analyses were performed using both MrBayes 3.2.2 (Ronquist et al., 2012) for Bayesian inference analyses and RaxML web-server (Stamatakis et al., 2008) and GARLI 2.0 (Zwickl, 2006) for maximum likelihood (ML). We first analyzed each gene separately using ML criteria to check for any contamination, unusual or conflicting results from sequencing. For individual gene analyses, we conducted ML using RaxML to find the tree with the best likelihood score. One hundred pseudo bootstrap replicates were also performed to assess branch support for each individual tree (supplementary materials). We subsequently concatenated genes into three datasets, two based on genome of origin (Nuclear or Mitochondrial), and one based on the full concatenated dataset before tree searches were conducted under Maximum Likelihood and Bayesian models (Fig. 2). For the mtDNA, nDNA and concatenated analyses, we used models and partition schemes from PartitionFinder and ran four independent Bayesian analyses in MrBayes each with one hot chain and three cold chains. Each run started with a random tree sampling every one thousand generations for ten million generations with a burnin of 10% using default priors except that the parameters statefreq, revmat, shape, and pinvar were unlinked between partitions. We used Tracer 1.5 (Rambaut et al., 2014) and AWTY (Wilgenbusch et al., 2004) to determine MCMC convergence for all Bayesian analysis. For the GARLI analysis, we used the same models and partition scheme as the Bayesian analysis and conducted ten ML tree searches with default settings using a random starting tree to find the tree with the best likelihood score. To assess branch support one thousand Maximum Likelihood bootstrap replicates were conducted in GARLI, Maximum likelihood bootstrap trees were summarized in Sumtrees ver. 3.1.0 (Sukumaran and Holder, 2010) with a minimum clade frequency of 50% and branch support was mapped onto the best scoring ML tree. Trees were visualized using FigTree v1.4.0 (Rambaut, 2009) and rooted with *Dichelia histrionana* (Tortricidae), since *Dichelia* is considered the closest genus to *Syndemis* in recent phylogenetic work (Dombroskie, 2011).

A haplotype network was used to better understand how variation in this genome reflected the patterns and levels of diversification from the evolution-based models. We used the program Popart (Leigh and Bryant, 2015) to construct the haplotype network using the mtDNA dataset and the TCS network method (Clement et al., 2002) We colorcoded taxa by geographic location/hostplant.

2.4. Hypothesis testing of tree topologies

Due to incongruence in genomes, specifically the question of a significant difference in genome-based topologies in the Central Coast California *Syndemis*, hypothesis testing of tree topologies was necessary. Two trees, one with a paraphyletic Central Coast California *Syndemis* (as reflected in the mtDNA [Fig 3]) and another with a monophyletic Central Coast California *Syndemis* (as reflected in the nDNA [Fig 3]) were tested using the SH test (Shimodaira and Hasegawa, 2001) implemented in the phangorn package (Schliep, 2011) in the program R. We used two datasets (mtDNA and nDNA) to test for significant differences between the two trees (eg monophyletic and paraphyletic Central Coast California *Syndemis* [Fig. 3]). We used optim.pml command in the phangorn package to calculate log-likelihood scores given a tree (Fig. 3; nDNA and mtDNA) and given a dataset (nDNA and mtDNA). We then con-

Table 1
Species, voucher numbers, collection data, DEC model range and Genebank accession numbers for the taxa used in this study.

Species	Voucher #	Collection locality	Range	COI	COII	COIII	CAD	EF1A	MDH	NAD1	RPS5
Dichelia histrionana	jd4556del	Denmark: LFM: Mandemarke, VII-27-28-2007 O. Karsholt		KY501293	KY501342		KY501586	KY501635	KY501439	KY501488	KY501537
Syndemis afflictana	as06koot	ID: Kootenai County, VI-9-2000 D. Rubinoff	East of the Rocky mountains	KY501272	KY501321	KY501370	KY501565	KY501614	KY501418	KY501467	KY501516
Syndemis afflictana	dr120albt	CAN: Alberta, Sherwood Park V-31-2001 G. Pohl	East of the Rocky mountains	KY501245	KY501294	KY501343	KY501538	KY501587	KY501391	KY501440	KY501489
Syndemis afflictana	dr126tenn	USA: TN: Blount Co. Great Smoky Mountains National Park V-22-2001 D. Rubinoff	East of the Rocky mountains	KY501251	KY501300	KY501349	KY501544	KY501593	KY501397	KY501446	KY501495
Syndemis afflictana	ms552albt	CAN: Alberta, Dunvegan V-10-2006: J. Dombroskie, D. Macaulay, S. Bromilow: MV Light	East of the Rocky mountains	KY501277	KY501326	KY501375	KY501570	KY501619	KY501423	KY501472	KY501521
Syndemis afflictana	ms564albt	CAN: Alberta, Buck Lake V-31-2007: M. Schwarzfeld: MV light	East of the Rocky mountains	KY501283	KY501332	KY501381	KY501576	KY501625	KY501429	KY501478	KY501527
Syndemis afflictana	dr127tenn	USA: TN: Blount Co. Great Smoky Mountains National Park V-22-2001 D. Rubinoff	East of the Rocky mountains	KY501252	KY501301	KY501350	KY501545	KY501594	KY501398	KY501447	KY501496
Syndemis musculana	as310dnmk	Denmark: LFM: Mandemarke 19-25 vii 2007 O. Karsholt	Old World	KY501276	KY501325	KY501374	KY501569	KY501618	KY501422	KY501471	KY501520
Syndemis musculana	ms561aus	Austria: Teiermark Turacher Hohr 785m IV-7-2009 P. Huemer	Old World	KY501281	KY501330	KY501379	KY501574	KY501623	KY501427	KY501476	KY501525
Syndemis musculana	ms562aus	Austria: Prov cuneo Entracque 7300m Cal della Povina X-6-2009 P. Huemer	Old World	KY501282	KY501331	KY501380	KY501575	KY501624	KY501428	KY501477	KY501526
Syndemis musculana	ms600rom	Romania: Transylvania: Apuseni Mountains V-19-2009 D. Rubinoff	Old World	KY501287	KY501336	KY501385	KY501580	KY501629	KY501433	KY501482	KY501531
Syndemis musculana	ms608rom	Romania: Transylvania: Sambata de sus, V-21-2009 D. Rubinoff	Old World	KY501288	KY501337	KY501386	KY501581	KY501630	KY501434	KY501483	KY501532
Syndemis sp.	dr122mcld	CA: Siskiyou Co. McCloud Area loop VI-8-2001 D. Rubinoff	Cascade Mountains	KY501247	KY501296	KY501345	KY501540	KY501589	KY501393	KY501442	KY501491
Syndemis sp.	dr123mcld	CA: Siskiyou Co. McCloud Area loop VI-8-2001 D. Rubinoff	Cascade Mountains	KY501248	KY501297	KY501346	KY501541	KY501590	KY501394	KY501443	KY501492
Syndemis sp.	dr124mcld	CA: Siskiyou Co. McCloud Area loop VI-8-2001 D. Rubinoff	Cascade Mountains	KY501249	KY501298	KY501347	KY501542	KY501591	KY501395	KY501444	KY501493
Syndemis sp.	dr138mcld	CA: Siskiyou Co. McCloud Area loop VI-8-2001 D. Rubinoff	Cascade Mountains	KY501261	KY501310	KY501359	KY501554	KY501603	KY501407	KY501456	KY501505
Syndemis sp.	cp25anvo	CA: San Mateo Co. Ano Nuevo State Park V-5-2002 J. Powell	Central California Coast	KY501273	KY501322	KY501371	KY501566	KY501615	KY501419	KY501468	KY501517
Syndemis sp.	cp26anvo	CA: San Mateo Co. Ano Nuevo State Park V-5-2002 J. Powell	Central California Coast	KY501274	KY501323	KY501372	KY501567	KY501616	KY501420	KY501469	KY501518
Syndemis sp.	cp27anvo	CA: San Mateo Co. Ano Nuevo State Park V-5-2002 J. Powell	Central California Coast	KY501275	KY501324	KY501373	KY501568	KY501617	KY501421	KY501470	KY501519
Syndemis sp.	dr130cam	CA: San Luis Obispo Co. Cambria Pines D. Rubinoff	Central California Coast	KY501255	KY501304	KY501353	KY501548	KY501597	KY501401	KY501450	KY501499
Syndemis sp.	dr131cam	CA: San Luis Obispo Co. Cambria Pines D. Rubinoff	Central California Coast	KY501256	KY501305	KY501354	KY501549	KY501598	KY501402	KY501451	KY501500
Syndemis sp.	dr134cam	CA: San Luis Obispo Co. Cambria Pines D. Rubinoff	Central California Coast	KY501257	KY501306	KY501355	KY501550	KY501599	KY501403	KY501452	KY501501
Syndemis sp.	dr135cam	CA: San Luis Obispo Co. Cambria Pines D. Rubinoff	Central California Coast	KY501258	KY501307	KY501356	KY501551	KY501600	KY501404	KY501453	KY501502
Syndemis sp.	dr150lobos	CA: Monterey Co. Pt. Lobos State Park X-15-01 D. Rubinoff	Central California Coast	KY501264	KY501313	KY501362	KY501557	KY501606	KY501410	KY501459	KY501508
Syndemis sp.	dr165anvo	CA: San Mateo Co. Ano Nuevo State Park V-5-2002 J. Powell	Central California Coast	KY501265	KY501314	KY501363	KY501558	KY501607	KY501411	KY501460	KY501509
Syndemis sp.	dr166anvo	CA: San Mateo Co. Ano Nuevo State Park V-5-2002 J. Powell	Central California Coast	KY501266	KY501315	KY501364	KY501559	KY501608	KY501412	KY501461	KY501510
Syndemis sp.	dr179mpine	CA: Monterey Co. Monterey Peninsula XI-10-11-2002 J. Powell	Central California Coast	KY501268	KY501317	KY501366	KY501561	KY501610	KY501414	KY501463	KY501512
Syndemis sp.	dr178mpine	CA: Monterey Co. Monterey Peninsula XI-10-11-2002 J. Powell	Central California Coast	KY501267	KY501316	KY501365	KY501560	KY501609	KY501413	KY501462	KY501511
Syndemis sp.	as01bcrk	USA :CA: Monterey County V-31-2002 J. Powell	Central California Coast	KY501269	KY501318	KY501367	KY501562	KY501611	KY501415	KY501464	KY501513

(*continued on next page*)

Table 1 (*continued*)

Species	Voucher #	Collection locality	Range	COI	COII	COIII	CAD	EF1A	MDH	NAD1	RPS5
Syndemis sp.	as02bcrk	USA :CA: Monterey County V-31-2002 J. Powell	Central California Coast	KY501270	KY501319	KY501368	KY501563	KY501612	KY501416	KY501465	KY501514
Syndemis sp.	dr121bcrk	USA: CA: Monterey Co. Big Creek V-31-2001 J. Powell	Central California Coast	KY501246	KY501295	KY501344	KY501539	KY501588	KY501392	KY501441	KY501490
Syndemis sp.	dr125bcrk	USA: CA: Monterey Co. Big Creek V-31-2001 J. Powell	Central California Coast	KY501250	KY501299	KY501348	KY501543	KY501592	KY501396	KY501445	KY501494
Syndemis sp.	dr136bcrk	USA: CA: Monterey Co. Big Creek V-31-2001 J. Powell	Central California Coast	KY501259	KY501308	KY501357	KY501552	KY501601	KY501405	KY501454	KY501503
Syndemis sp.	dr137bcrk	USA: CA: Monterey Co. Big Creek V-31-2001 J. Powell	Central California Coast	KY501260	KY501309	KY501358	KY501553	KY501602	KY501406	KY501455	KY501504
Syndemis sp.	ms643bcrk	USA: Ca: Monterey Co. Big Creek Reserve 125m V-31-02 J. Powell D. Rubinoff	Central California Coast	KY501289	KY501338	KY501387	KY501582	KY501631	KY501435	KY501484	KY501533
Syndemis sp.	ms644bcrk	USA: Ca: Monterey Co. Big Creek Reserve 125m V-31-02 J. Powell D. Rubinoff	Central California Coast	KY501290	KY501339	KY501388	KY501583	KY501632	KY501436	KY501485	KY501534
Syndemis sp.	ms645bcrk	USA: Ca: Monterey Co. Big Creek Reserve 125m V-31-02 J. Powell D. Rubinoff	Central California Coast	KY501291	KY501340	KY501389	KY501584	KY501633	KY501437	KY501486	KY501535
Syndemis sp.	ms646bcrk	USA: Ca: Monterey Co. Big Creek Reserve 125m V-31-02 J. Powell D. Rubinoff	Central California Coast	KY501292	KY501341	KY501390	KY501585	KY501634	KY501438	KY501487	KY501536
Syndemis sp.	dr148hum	Ca: Humboldt Co. Kneeland V-30-2001 R. Wielgus	Northern California Coast	KY501263	KY501312	KY501361	KY501556	KY501605	KY501409	KY501458	KY501507
Syndemis sp.	ms554sono	Ca: Sonoma Co. West of Plantation VII-9-28-09 Ret-c Lure J. Powell	Northern California Coast	KY501278	KY501327	KY501376	KY501571	KY501620	KY501424	KY501473	KY501522
Syndemis sp.	ms555sono	Ca: Sonoma Co. West of Plantation VII-9-28-09 Ret-c Lure J. Powell	Northern California Coast	KY501279	KY501328	KY501377	KY501572	KY501621	KY501425	KY501474	KY501523
Syndemis sp.	ms557sono	Ca: Sonoma Co. West of Plantation VII-9-28-09 Ret-c Lure J. Powell	Northern California Coast	KY501280	KY501329	KY501378	KY501573	KY501622	KY501426	KY501475	KY501524
Syndemis sp.	ms570hum	Ca: Humboldt Co. Kneeland 17 June 2009 Sticky trap pheromone lure R. Wielgus	Northern California Coast	KY501284	KY501333	KY501382	KY501577	KY501626	KY501430	KY501479	KY501528
Syndemis sp.	ms591hum	Ca: Humboldt Co. Kneeland 02 July 2009 Sticky trap pheromone lure R. Wielgus	Northern California Coast	KY501285	KY501334	KY501383	KY501578	KY501627	KY501431	KY501480	KY501529
Syndemis sp.	ms592hum	Ca: Humboldt Co. Kneeland 02 July 2009 Sticky trap pheromone lure R. Wielgus	Northern California Coast	KY501286	KY501335	KY501384	KY501579	KY501628	KY501432	KY501481	KY501530
Syndemis sp.	as03sier	CA: Sierra County Yuba Pass VI-29-2001 J. Powell	Sierra Nevada	KY501271	KY501320	KY501369	KY501564	KY501613	KY501417	KY501466	KY501515
Syndemis sp.	dr128sier	CA: Sierra Co. Yuba Pass VI-29-2001 J. Powell	Sierra Nevada	KY501253	KY501302	KY501351	KY501546	KY501595	KY501399	KY501448	KY501497
Syndemis sp.	dr129sier	CA: Sierra Co. Yuba Pass VI-29-2001 J. Powell	Sierra Nevada	KY501254	KY501303	KY501352	KY501547	KY501596	KY501400	KY501449	KY501498
Syndemis sp.	dr139sier	CA: Sierra Co. San Francisco State Field Station VI-28-2001 J. Powell	Sierra Nevada	KY501262	KY501311	KY501360	KY501555	KY501604	KY501408	KY501457	KY501506

Table 2
Primers names, sequences and annealing temperatures utilized in this study.

Gene	Primer name	Direction	Sequence (5′–3′)	Published	Annealing temp (°C)
COI	LCO-1490	Forward	GCTCAACAAATCATAAAGATATTGG	Folmer et al. (1994)	50
	HCO-2198	Reverse	TAAACTTCAGGGTGACCAAAAAATCA	Folmer et al. (1994)	
	Jerry	Forward	CAACATTTATTTTGATTTTTTGG	Simon et al.(1994)	50
	Pat	Reverse	TCCAATGCACTAATCTGCCATATTA	Simon et al. (1994)	
COII	TK-N3796	Forward	ACTATAAAATGGTTTAAGAG	Simon et al. (2006)	50
	TL2-J3043	Reverse	GGCAGACTATATGYAATGRATTTAA	Simon et al. (2006)	
COIII	C3-J4792	Forward	GTAGATTATAGACCWTGRCC	Simon et al. (2006)	50
	C3-N5460	Reverse	TCTACAAAATGTCARTAYCA	Simon et al. (2006)	
NAD1	N1-J11876	Forward	CGAGGTAAAGTMCCWCGAACYCA	Simon et al. (2006)	60
	N1-N12595	Reverse	GTWGCTTTTTTGACTTTATTRGARCG	Simon et al. (2006)	
28s-rDNA	28s-III-F	Forward	CCGCTAAGGAGTGTGTAA	Hillis and Dixon (1991)	55
	28s-III-R	Reverse	GAAGTTACGGATCTARTTTG	Hillis and Dixon (1991)	
CAD	Syn	Forward	CCGGCAAGTCCACCACCACCGGCC	Designed for this study	57
	Demis	Reverse	GAAGTCGGCGCCCTTGGGCGGG	Designed for this study	
EF-1a	Mari	Forward	GGGATTTGGCTAAATTTACAAGGG	Designed for this study	55
	Posa	Reverse	GGTCCCTAAGATCTTCGC	Designed for this study	
MDH	Syn-MDH-F	Forward	ATGATGGGGGTTCTGGAGGGTG	Designed for this study	60
	Syn-MDH-R	Reverse	AACAGGGAATGAGAAGACAACG	Designed for this study	
RPS5	Syn-RPS5-F	Forward	TGGAACGACGACAGCGTGGAGGCGGG	Designed for this study	55
	Syn-RPS5-R	Reverse	GCAACACGTTCCAGCTCATCCTTC	Designed for this study	

Table 3
Substitution model and partition selected by Partition Finder.

Gene/codon position	Partition Finder Partition	Substitution model Partition Finder
COI_1	2	GTR+I+G
COI_2	6	GTR+I+G
COI_3	4	GTR+I+G
COII_1	4	GTR+I+G
COII_2	6	GTR+I+G
COII_3	4	GTR+I+G
COIII_1	4	GTR+I+G
COIII_2	4	GTR+I+G
COIII_3	5	GTR+G
NADH1_1	5	GTR+G
NADH1_2	4	GTR+I+G
NADH1_3	2	GTR+I+G
CAD_1	4	GTR+I+G
CAD_2	2	GTR+I+G
CAD_3	4	GTR+I+G
EF1a_1	2	GTR+I+G
EF1a_2	3	GTR+I+G
EF1a_3	1	GTR+G
MDH_1	3	GTR+I+G
MDH_2	1	GTR+G
MDH_3	2	GTR+I+G
RPS5_1	1	GTR+G
RPS5_2	2	GTR+I+G
RPS5_3	3	GTR+I+G

ducted the SH-test using the SH.test command using the two tree topologies and calculated log-likelihood scores and tested for significance using 10,000 bootstrap replicates (Table 5).

2.4.1. Species tree analysis

Species tree analyses were implemented using ∗BEAST as part of the BEAST 1.8 package using the concatenated, mtDNA and nDNA datasets (8, 4 and 4 genes respectively) and number of gene copies per lineage ranged from one (outgroup) to 12 (average non-outgroup gene copies = 6.86). ∗BEAST infers species trees from multilocus data using the multispecies coalescent model. The substitution model for each gene was selected from the best model found using jModeltest (Table 4). We did not use the partition scheme and model given by PartitionFinder for these analyses because the best-fit partition scheme the program gave combined individual codon positions from nuclear and mitochondrial genomes in a single partition. Additionally, we were unable to run species tree analyses using PartitionFinder due to low tree heights in some partitions caused by low variation in some of the genes we used. Species were designated based on species' range as derived from our collection records and those in museum collections. For most species there is a geographic gap in distribution such that they could be defined by contiguous physical locations with the exception of the central California coast species that were split into two taxa based on larval hostplant use (redwood vs. Monterey pine feeding species). We used a strict molecular clock model with normally distributed rate variation for all genes. A Yule process tree prior and piece-wise linear and constant root population prior was used with the ploidy of each gene set to the genome of origin. For each dataset, we ran four independent chains for 100 million generations sampling every 10,000 generations and discarded the first 10,000,000 as burnin. We combined the results of the four separate runs and visualized them in Tracer

Table 4
Genetic variation and substitution model from jModeltest selected for each gene.

Gene	Parsimony informative	Variable	Total Characters	Substitution Model jModeltest
COI	95 (6.55%)	101 (6.96%)	1451	HKY+G
COII	46 (6.31%)	50 (6.86%)	729	HKY+G
COIII	44 (6.74%)	50 (7.66%)	653	HKY+G
NADH1	50 (8.55%)	53 (9.06%)	585	HKY+G
Cad	19 (2.44%)	19 (2.44%)	779	HKY+G
EF1a	10 (1.34%)	20 (2.68%)	745	HKY
MDH	22 (3.32%)	26 (3.93%)	662	K80+G
RPS5	5 (0.80%)	5 (0.80%)	628	HKY
mtDNA	235 (6.88%)	254 (7.43%)	3418	
Nuclear DNA	56 (1.59%)	51 (1.45%)	2814	
All Genes	291 (4.19%)	305 (4.40%)	6232	

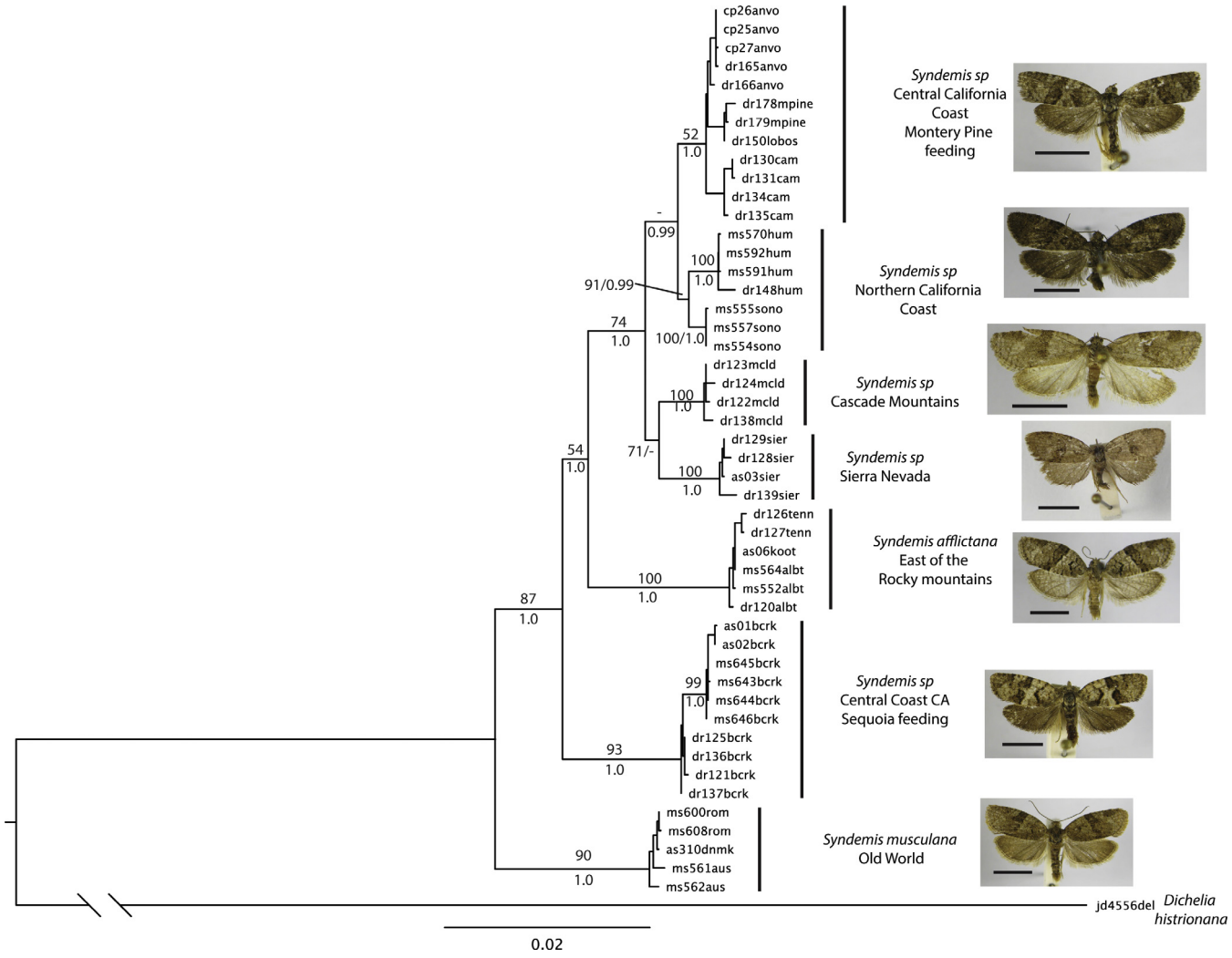

Fig. 2. Maximum likelihood tree, concatenated dataset. Support values above branches are Maximum Likelihood Bootstrap values and below are Bayesian Posterior Probabilities. Scale bar indicates the number of substitutions per site. Big Creek population is basal within the tree similar to the mtDNA dataset.

1.5 to check for convergence and to confirm effective sample size of the parameter estimates. We used TreeAnnotator 1.8 to obtain Posterior probabilities for the species tree, and FigTree to examine the maximum clade credibility tree. We then used DensiTree (Bouckaert, 2010) to visualize all the topologies included in the posterior distribution of trees for each dataset.

2.5. Biogeography

Due to the dramatic incongruence in genomes regarding the placement of the central coast redwood-feeding populations, we conducted separate biogeographic analyses. We used the best scoring ML tree from the mitochondrial, nuclear and concatenated datasets to reconstruct the ancestral geographic range for *Syndemis* using two packages; Biogeobears (Matzke, 2013) and RASP 2.1b (Yu et al., 2011). We estimated a dispersal-extinction-cladogenesis (DEC) model and a DEC+jump (DEC+J) dispersal model using Biogeobears (Ree et al., 2005; Ree and Smith, 2008; Matzke, 2014) to infer the ancestral geographic range for *Syndemis*. The DEC and DEC+J models are continuous time stochastic models for geographic range evolution in discrete areas. Maximum likelihood parameters are used to estimate historical rates of dispersal and extinction within areas. We modified our ML trees to scale them to absolute time using the ape package (Paradis et al., 2004) in program R, and edited the tree topology by removing the outgroup and redundant samples to accommodate a requirement of Biogeobears for the tree to be completely bifurcating and to have branch lengths between nodes greater than zero. We constructed DEC and DEC+J models using the present-day geographic ranges for each population; each OTU in the phylogeny was assigned to one region based on present-day distributions (Table 1; Fig. 1). We then compared log likelihood scores of each model using a chi-squared test in R to test if the models were significantly different from each other.

We also used the program RASP 2.1b (Yu et al., 2011) for ancestral biogeographic reconstruction. We used the Bayesian ancestral state reconstruction method implemented in RASP that determines the probability of each ancestral area averaged over all sampled trees in the posterior distribution. We coded each taxon according to biogeographic region applying an equal rates model for each geographic area. We ran the analysis using 10 chains for 5 million generations in RASP, sampling every 1000th generation with a burnin of 2 million generations.

3. Results

3.1. Morphological data

The wings of North Coast, Central California Monterey pine and redwood-feeding populations were significantly darker, especially

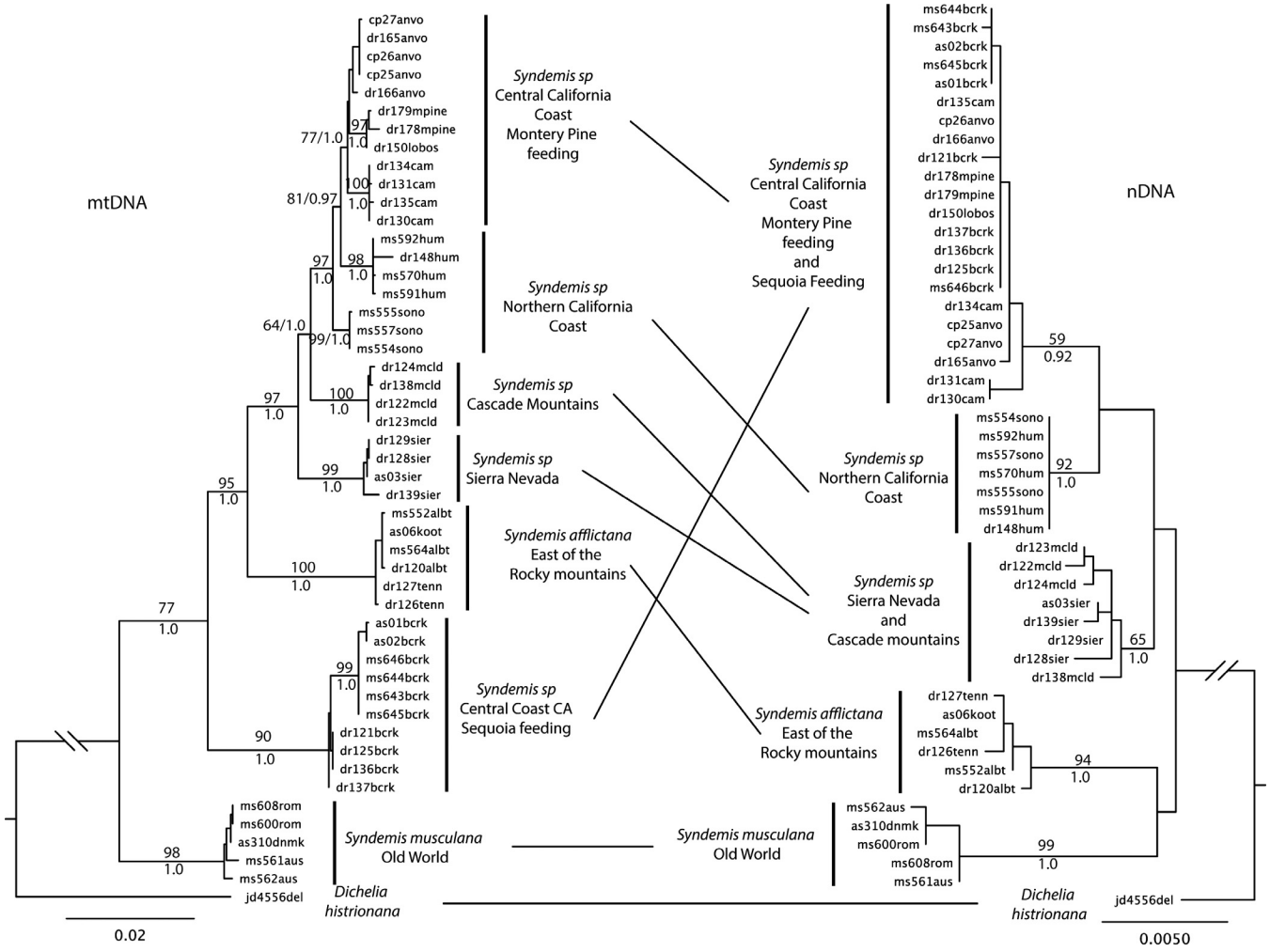

Fig. 3. Maximum likelihood trees, mtDNA and nDNA datasets. Tanglegram shows changes in topology between mtDNA and nDNA datasets. Support values above branches are Maximum Likelihood Bootstrap values and below are Bayesian Posterior Probabilities. Only Maximum Likelihood bootstrap values above 50 and Bayesian Posterior probabilities above 0.9 are shown. Scale bar under trees indicates the number of substitutions per site. Note Sequoia feeding population is a basal branching clade within the mtDNA tree however it is mixed with the Monterey pine feeding population in the nDNA tree.

Table 5
SH values for the mitochondrial and nuclear phylogenetic trees.

Sequence data	Tree	OBS lnL	Difference (lnL)	P-value
Mitochondrial	Nuclear	−11791.604	3633.18	<0.00001
Mitochondrial	Mitochondrial	−8158.424	0	0.4958
Nuclear	Nuclear	−5173.954	0	0.4825
Nuclear	Mitochondrial	−5374.704	200.7495	<0.00001

on the hindwings, than any other populations of *Syndemis* across North America and Europe (Figs. 1 and 2). Other populations shared a very similar color and pattern scheme. However, extensive dissections revealed no reliable characters in the genitalia of either sex that could be used diagnostically (Powell and Rubinoff, unpublished). This is surprising since genitalic characters are the standard method of discerning species in Lepidoptera (eg Horak, 1999). Because morphology was not phylogenetically informative in this study, it is not considered further.

3.2. Phylogenetic analysis

Trees constructed using mtDNA, nDNA and concatenated, are shown in Figs. 2 and 3. There is conflict between genomes (mtDNA vs. nDNA) regarding the phylogenetic placement of the Central California taxa. Mitochondrial DNA data indicate a highly resolved tree with the Central Coast California *Syndemis* paraphyletic. The Big Creek redwood-feeding population is sister to a clade comprising all other North American populations with mtDNA. However, the nuclear data resulted in a less resolved tree, with far fewer informative characters and variable sites (Table 4), and a monophyletic Central Coast California *Syndemis* with the Big Creek redwood-feeding population indistinguishable from the *P. radiata* Central coast populations. Incongruence between genomes is often thought to be the result of mtDNA introgression (Funk and Omland, 2003), yet when mtDNA and nDNA were concatenated, ML and Bayesian analysis produced trees similar in topology to the mtDNA analyses.

3.3. Haplotype network

The haplotype network (Fig. 4) reflects the same patterns of isolation found from the phylogenetic analysis of mtDNA data. Most

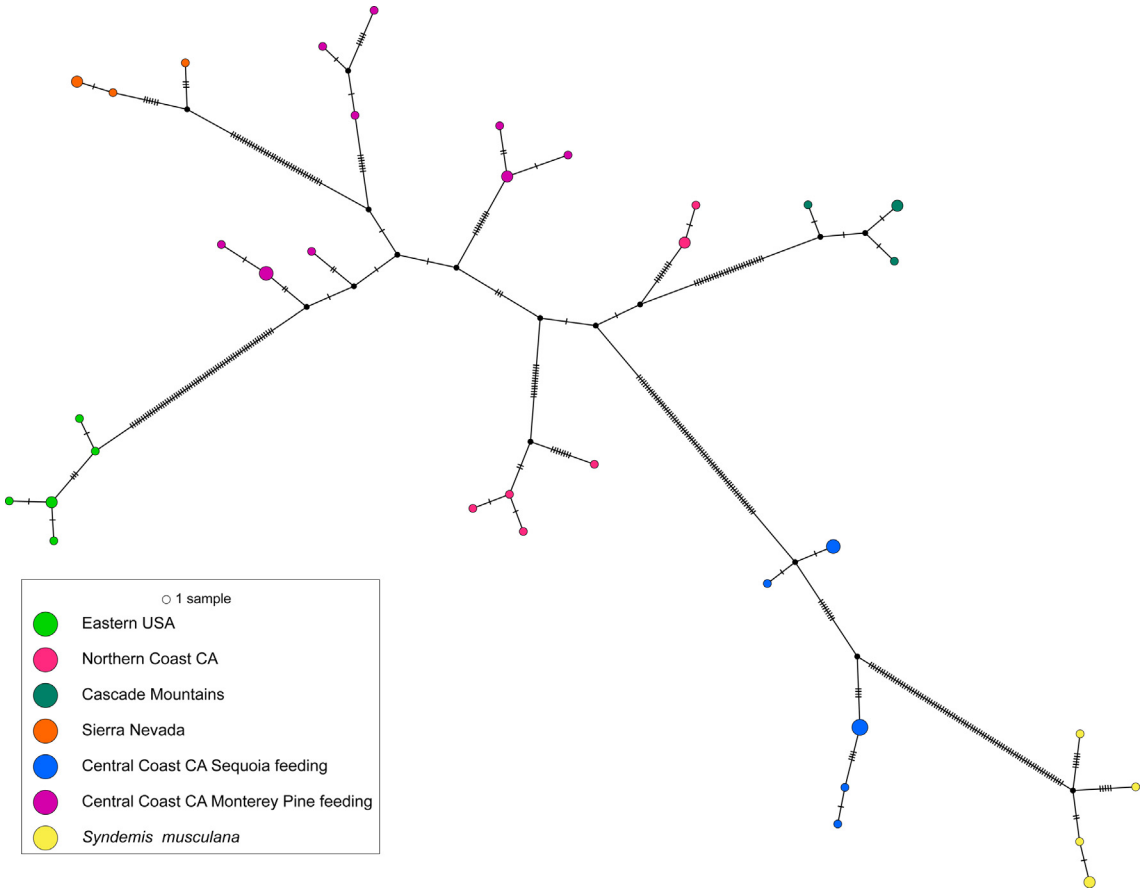

Fig. 4. Mitochondrial Haplotype network of *Syndemis* populations constructed using Popart (Leigh and Bryant, 2015). Similar to the phylogenetic analyses, the haplotype network shows high divergence between species/populations. Note the large divergences found in Central Coast CA between the *Sequoia* and Monterey Pine feeding populations.

Table 6
Biogeographic model comparison DEC vs DEC+J.

Dataset	Model	log-Likelihood	P-value
mtDNA	DEC	−18.0105	1
mtDNA	DEC+J	−18.011	
nDNA	DEC	−20.4258	**0.002**
nDNA	DEC+J	−15.6534	
Concatinated	DEC	−20.6097	**0.023**
Concatinated	DEC+J	−18.0146	

Bold indicates significant p-value.

populations are clearly monophyletic with large numbers of base pair changes isolating them from their nearest neighbor. The Monterey pine feeding populations have the greatest amount of haplotype variation, with northern California coastal populations being the next most variable, despite many other populations having been sampled from across far larger geographic regions (eg eastern 2/3 of North America, or across Europe).

3.4. Hypothesis testing

Results from the SH-test are shown in Table 5. Results indicate that the mtDNA supports a significantly different topology and placement of the Central Coast California *Syndemis* from the nDNA, which has all Central Coast California *Syndemis* being monophyletic, with redwood and Monterey pine feeding populations as sister taxa. Also the difference in log-likelihood scores between the two topologies when based upon the nDNA dataset vs mtDNA data set was significantly smaller (200[nDNA] vs 3633[mtDNA]).

3.4.1. Species tree analysis

Species trees using three different datasets corresponding to genome of origin are shown in Figs. 5–7 (mtDNA, nDNA, and concatenated). Species tree and Cloudogram of the mtDNA dataset shows high support for the basal redwood-feeding population however the nDNA dataset shows the same redwood- feeding population together with the other California populations. Concatenated species tree analysis results in a topology congruent with the mtDNA analysis but reflecting some uncertainty in the topology with regards to the redwood-feeding populations and the Palearctic *S. musculana*.

3.5. Biogeography

Biogeographic analysis on three ML trees obtained from the mtDNA, nDNA and concatenated datasets are shown in Figs. 8–10. Using the adjacency matrix and the DEC and DEC+J models shown in Fig. 1, we used Biogeobears and RASP 2.1b to estimate the likelihood of the ancestral geographic range for *Syndemis*. Because our study focused on North American taxa, our sampling of Old World species was limited and insufficient to determine whether *Syndemis* originated in the Old or New World.

Biogeographic analysis support different biogeographic scenarios depending on the model, program and dataset used. Likelihood ratio tests of different models used for the Biogeobears results also supported different models depending on the dataset (Table 6). The

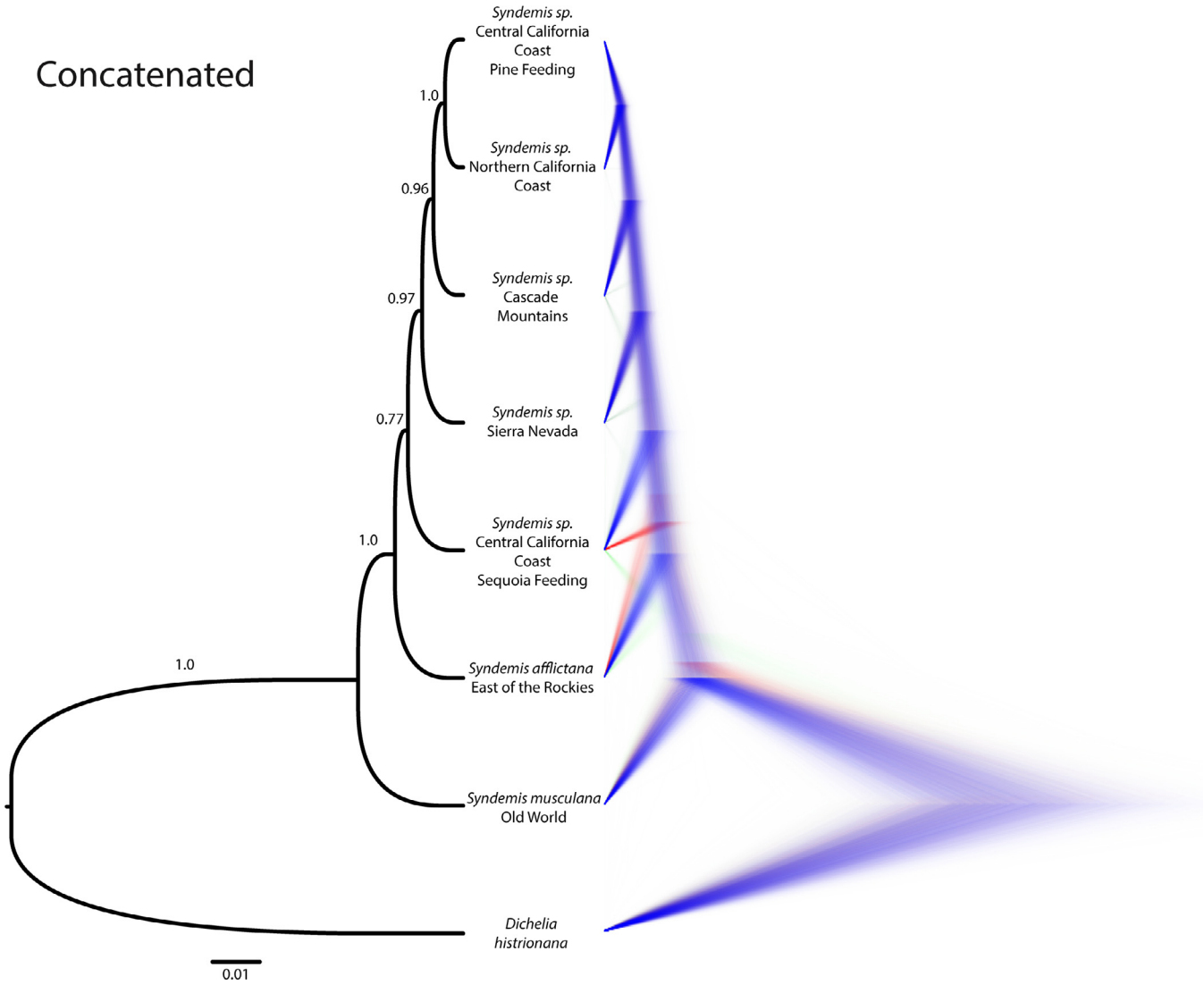

Fig. 5. Species tree (left) and cloudogram (right) of species tree analysis using concatenated dataset. Most popular tree topology is in blue, second most popular tree topology is in red and third most popular tree topology is green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

mitochondrial genome (Figs. 8–10, B) shows no difference between the DEC and the DEC+J models in Biogeobears and results in a vicariance like biogeographic scenario with the most likely MRCA biogeographic state encompassing all areas. However the nuclear genome favors the DEC+J which includes dispersal speciation events. In this analysis the most likely MRCA biogeographic state is within California. RASP analysis gives a California origin for both the mtDNA and the concatenated datasets with dispersal to eastern North America and then secondary back colonization returning California. The nuclear dataset indicates an east coast origin for the North American *Syndemis* which then dispersed to California.

4. Discussion

Only the mtDNA found in redwood-feeding populations at Big Creek is basal to the rest of the North American mtDNA haplotypes, suggesting much more recent, secondary, contact between redwood and adjacent Monterey pine-feeding populations. This pattern has been seen in other organisms in which females are philopatric or have strong host associations and males actively disperse, including Macaques (Tosi et al., 2000), and even other Lepidoptera (Ohshima and Yoshizawa, 2010). Due to female host specificity, populations with introgressed nDNA remain distinct, and genetic swamping does not occur due to extremely sex-biased hybridization. Thus, the nuclear and mitochondrial DNA produced trees with very different tree topologies. Mitochondrial data places the central coast redwood-feeding population of *Syndemis* as earlier branching than all other North American taxa, while nuclear data mixes it with the other Central California coast populations as a more recently derived apical branch. Our results support the already established caution that the use of mtDNA alone in systematics may be inadequate (eg Cognato, 2006; Rubinoff et al., 2006). Most often, mtDNA- nDNA conflict across species is given as an example of how mtDNA can be misleading, most often due to incomplete lineage sorting or secondary introgression, though the discordance can have many other causes (Ballard and Whitlock, 2004; Rubinoff and Holland, 2005; Rubinoff et al., 2006). While incomplete lineage sorting is often invoked to explain mtDNA-nDNA conflict, other phenomena, like hybridization, may be operating (Wahlberg et al., 2009). In the case of *Syndemis* neither seems a good explanation; incomplete lineage sorting should confuse resolution between more recently-related taxa where genomes have not had enough time to sort unequivocally, and would not be expected to bring a single apically branch-

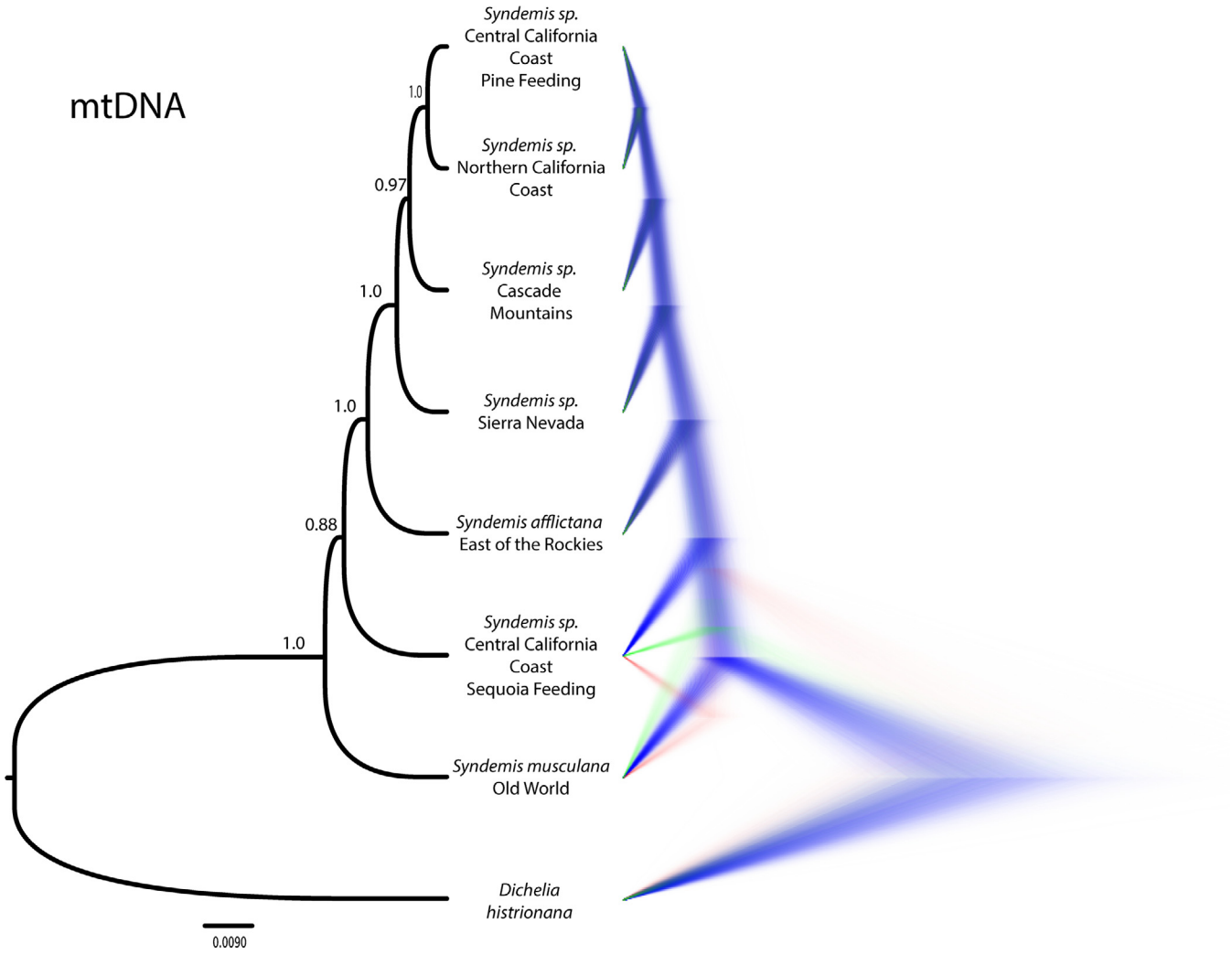

Fig. 6. Species tree (left) and cloudogram (right) of species tree analysis using mitochondrial dataset. Most popular tree topology is in blue, second most popular tree topology is in red and third most popular tree topology is green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ing species to a basal position relative to almost all other congeners. Hybridization might change sister taxon relationships, but in the case of the redwood-feeders, mtDNA renders them basally branching relative to the rest of the North American species, without a clear sister taxon based on their mtDNA, an independence which does not suggest a hybrid origin for the species. Importantly, it is usually the mtDNA, not the nDNA, introgressing that is blamed for the conflict (eg Funk and Omland, 2003). However, our results indicate that, depending on the ecology of the organisms in question, mtDNA can be the genome that *doesn't* introgress, and this phenomenon, and the resultant interpretation of a taxon's evolutionary history, merits a more careful consideration of perspective. Certainly, mtDNA-nDNA conflict is not uncommon, but in cases like *Syndemis*, the evolutionary history we might be most interested in understanding is wholly based on the genome that is considered most likely to be misleading. In taxa like *Syndemis*, natural history probably plays a pivotal role in this switch, but because of their small size, and very short lifespans, the non-migratory movements of most moth species in natural systems are virtually unknown. This work presents one of the first opportunities to understand, indirectly, how the different sexes behave over longer time scales.

While it is well-known that nDNA evolves more slowly than mtDNA and may cause discordance between genomic phylogenies (eg Kodandaramaiah et al., 2013; Hendrix et al., 2014) the conflict between mtDNA and nDNA in our data is most extreme concerning the Redwood feeding populations where the mtDNA dramatically changes their position. This suggests a conflict in phylogenetic signal between mtDNA and nDNA disproportionately focused on the Redwood feeding populations (Figs. 3 and 5–7) and likely reflects evolutionary history and male biased dispersal, though it could represent an artifact of differential divergence rates between genes (eg Peters et al., 2012). However if the discordance was the result of mtDNA evolving at a faster rate, we should expect such conflict to occur at a similar level elsewhere in *Syndemis* populations across North America or Europe. The fact that the topologies for all other *Syndemis* taxa are much more similar between the mtDNA and nDNA (different pairing of taxa, but similar order of taxon branching) suggests that something beyond differential rates is operating. If we are mistaken, and the discordance we observe only in redwood-feeding populations is the result of nDNA evolving more slowly than mtDNA or ecological traits, this more rapid mtDNA divergence, and ecological divergence, is largely concentrated in the redwood feeding species, to the exclusion of even neighboring Monterey-pine feeding species, which are changing at the same rate as the rest of the genus. This would be difficult to explain given the current adjacent locations of the central California populations (since the premise would be a recent divergence rather

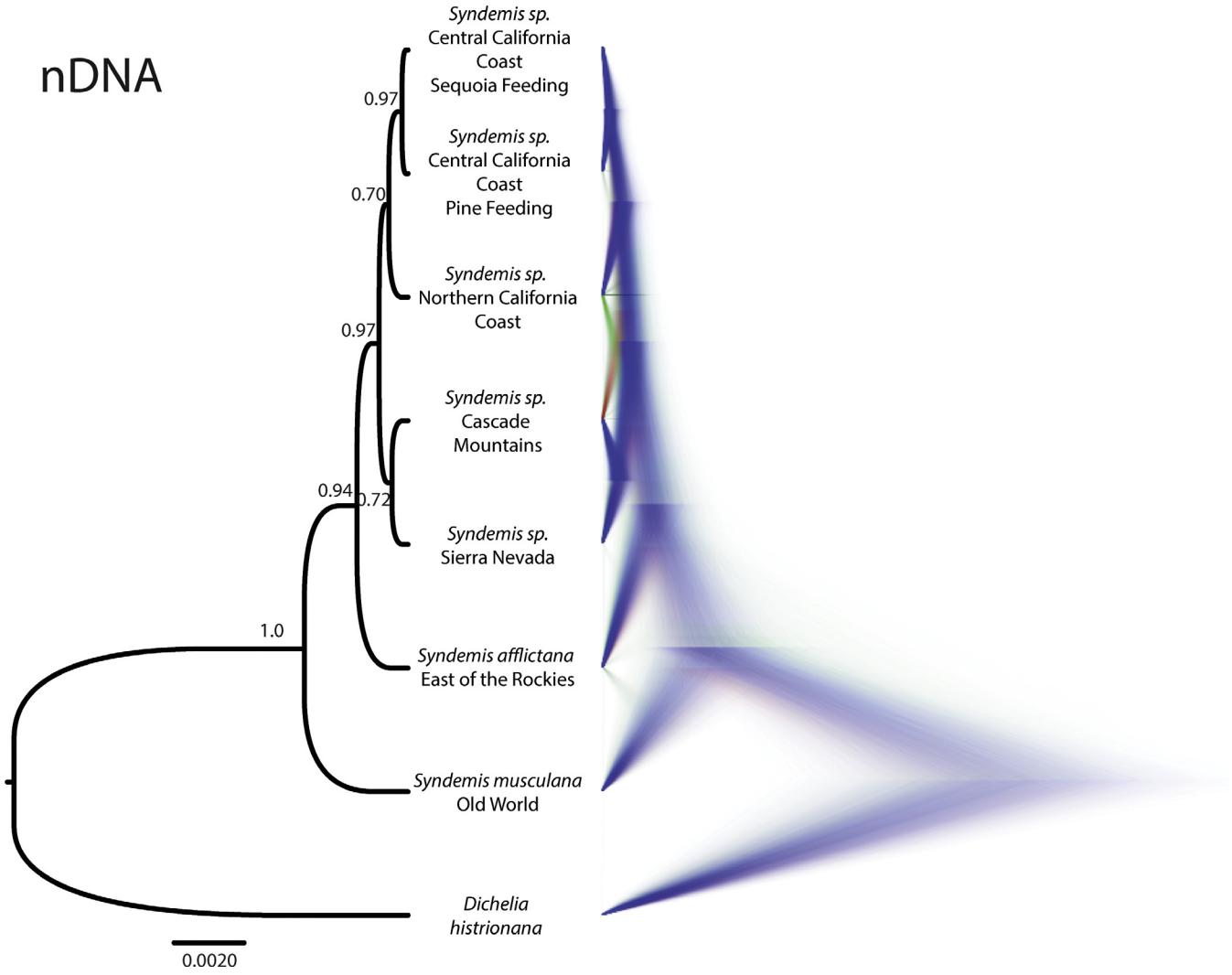

Fig. 7. Species tree (left) and cloudogram (right) of Species tree analysis using nuclear dataset. Most popular tree topology is in blue, second most popular tree topology is in red and third most popular tree topology is green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

than the secondary contact we propose). It would require the interspersed populations of the redwood feeding species to have diverged in concert in their mtDNA while each population was surrounded by Monterey Pine feeding populations, using the same pheromone, which retained very different mtDNA. While this is a possible explanation, it is less likely than our proposal that the Redwood feeding populations' mtDNA/nDNA discordance represents a much earlier invasion, divergence of mtDNA and secondary contact with male biased dispersal.

Despite trapping hundreds of male moths in adjacent Monterey pine and Redwood -feeding populations, we never saw the two species sympatrically. Yet, indirect evidence from our molecular data suggests that male *Syndemis* moths disperse more widely than females, mixing the nDNA between geographically proximate populations. Male moths are not bound to specific host plants or locations; in *Syndemis*, nDNA is not diagnostic for many, otherwise divergent, populations. Depending on population size and neutral selection, with as little as one immigrant individual every other generation, differentiation of populations due to genetic drift would be prevented (Slatkin, 1985). Thus, our results suggest there may be selection for retention of host preference and flight period such that even under gene flow from outside males which is sufficient to swamp nDNA differences in several genes, the differences in these ecological traits are retained in the redwood and Monterey pine feeding populations of Central California. Alternatively, it seems likely that these genes are either maternally inherited through mtDNA, or on the Y chromosome, since females are heterogametic in Lepidoptera. This possibility bears further examination in *Syndemis* and other taxa, since it may have important implications for speciation and host use across the order. There is growing evidence that mtDNA in animals is functionally important beyond just energy production and, in some cases, may have significant impacts on speciation (Breton et al., 2014).

The mitochondrial genetic diversity of *Syndemis* taxa in California far exceeds that of the rest of North America or Europe (California 1.6% vs. Eastern United States 0.2% vs. Europe 0.3%). Specifically, there is more variation across 50 miles of the central California coast than across the eastern 2/3 of North America or from Denmark to Romania. It is perhaps surprising that the greatest haplotype diversity was found in Monterey pine feeding populations (Fig. 4), considering they are one of the most geographically restricted. When contrasted with the lack of genetic variation between Alberta, Canada and Tennessee, or between Denmark and Romania, in *S. afflictana* and *S. musculana* respectively, the extreme variability in coastal California is all the more striking. The haplotype network supports the designation of the populations/species identified by the combined phylogenetic analysis, the taxonomy of which will be considered in a more specialized paper.

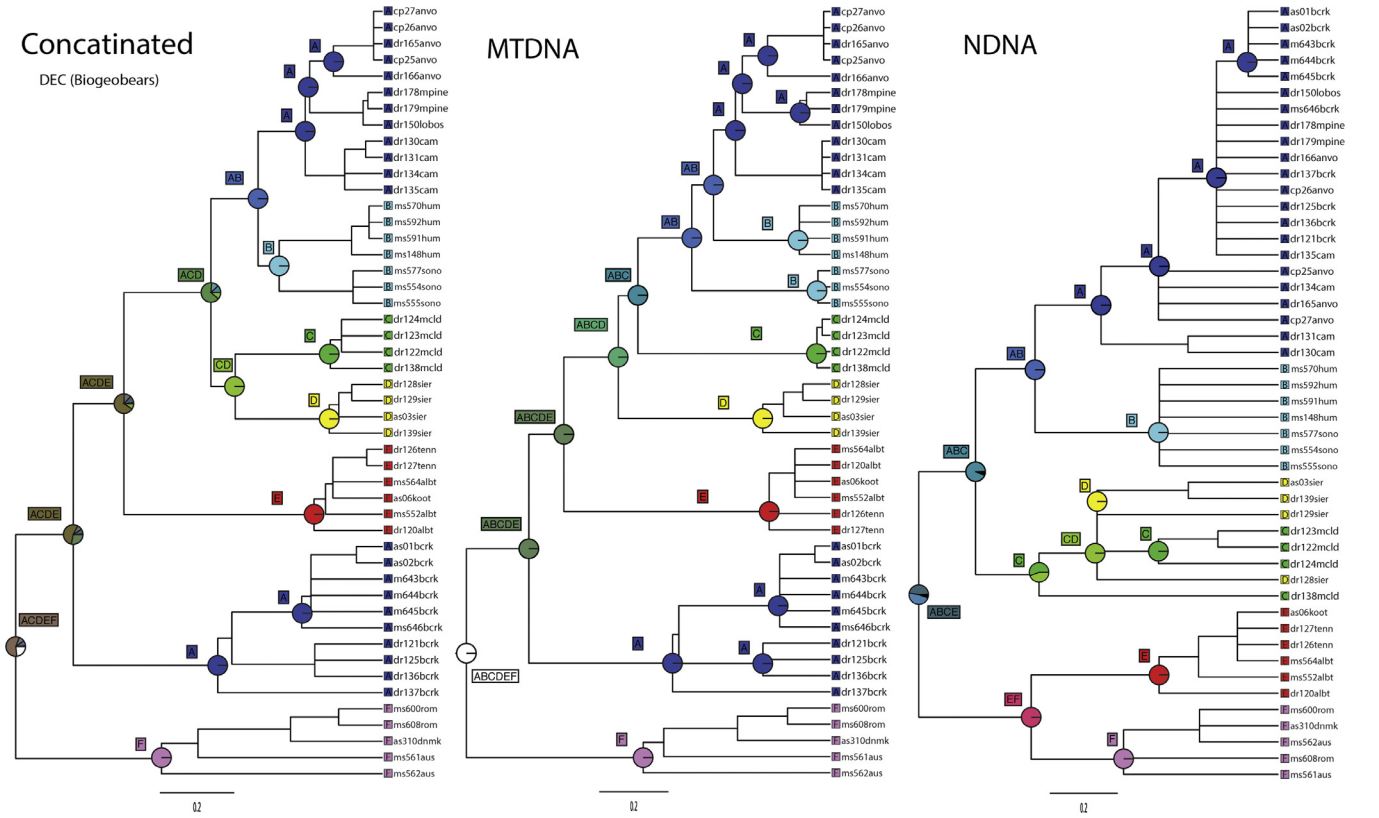

Fig. 8. Maximum likelihood trees, Likelihood analysis of ancestral geographic range using DEC model implemented in Biogeobears for all datasets (Concatinated, mtDNA and nDNA). Pie graphs on node indicated relative probability of ancestral geographic range colors and letters correspond to ranges indicated in map (Fig. 1). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Although genetic diversity between different taxa remained relatively low ⩽3.1% between Europe and California, the geographic distribution of the divergence was almost entirely centered in the California region, and particularly within the mtDNA. While the hosts for Sierran, Cascade and Humboldt county *Syndemis* are not known, they are likely to be conifers, based on the habitats of the moths, hosts of other, known, populations, and associated tree species in these areas. Given the earlier origin of the Big Creek redwood-feeding species' mtDNA and the unusual host use of both the central California species, it is likely that California was an ancient, if not the original region for the North American *Syndemis* taxa in this study. Using multiple methods (Matzke, 2013; Ree and Smith, 2008; Yu et al., 2011) to reconstruct the ancestral geographic range of *Syndemis* suggests that the Big Creek area in California is the most likely ancestral area for North American *Syndemis*. However, since redwood formerly had a much broader distribution across the American West, it is possible that redwood was an ancestral (but perhaps not exclusive) host for early *Syndemis*, and that the widespread, polyphagous habits of the species, which occur across much of central and eastern North America, are either more derived states or a retained ancestral polyphagy lost by redwood and Monterey pine feeding taxa. The Monterey pine feeding species is much more recently derived. This corresponds with the closed-cone pines being a more recent arrival in California (Axelrod, 1967), the ancestors of which evolved in Mexico and dispersed gradually up the west coast. While closed-cone pines, such as Monterey pine, are thought to have moved northward during the Miocene/Pliocene, redwoods were already widespread in Western North America at that time. It is likely that *Syndemis* did not encounter the closed-cone pines until they were sympatric with the moths in California (see below). Because there are other, putative, *Syndemis* species from Asia and Africa that we could not include, the historical biogeography of *Syndemis* may produce additional, unexpected, results as more taxa are added. However, the extremely localized diversity of Californian *Syndemis*, and the nature of the mtDNA and nDNA divergence are strongly supported and surprising results.

Mitochondrial DNA incongruence may reflect secondary nuclear introgression of the two coastal species, based on male cross attraction to female pheromones. The Monterey pine and redwood central coast species occur within 50 miles of each other, but are not currently sympatric. The localized ranges and apparent diversification onto different host plants in California stand in contrast to the very widespread and polyphagous *S. afflictana* and *S. musculana* and support the idea that host plant specialization may contribute to speciation, though the number of specialist species is too small to offer statistical power. *Syndemis* in California has more genetic divergence, host specialization and life cycle diversity than the taxa that occur across the eastern two-thirds of North America, or all of Europe. This pattern suggests that specialization and speciation have occurred in association with isolated, highly restricted host plants, even after they were relictual (as Monterey pine likely was by the time *Syndemis* colonized it). The early-branching central coast redwood-feeder mtDNA suggests that some ancient divergences (in this case in nDNA) may be swept away by genetic introgression from secondary invasions of related species. While mtDNA sweeps have been invoked to explain mtDNA introgression between species (Hurst and Jiggins, 2005), our data suggests the opposite; here is a situation where mtDNA is preserving an ancient divergence that has been largely obliterated by more recent gene flow in nDNA.

Disagreement between mtDNA and nDNA is often seen as a sign of introgression across mtDNA which obscures the true evolutionary history rendering mtDNA 'misleading' (Ballard and Whitlock,

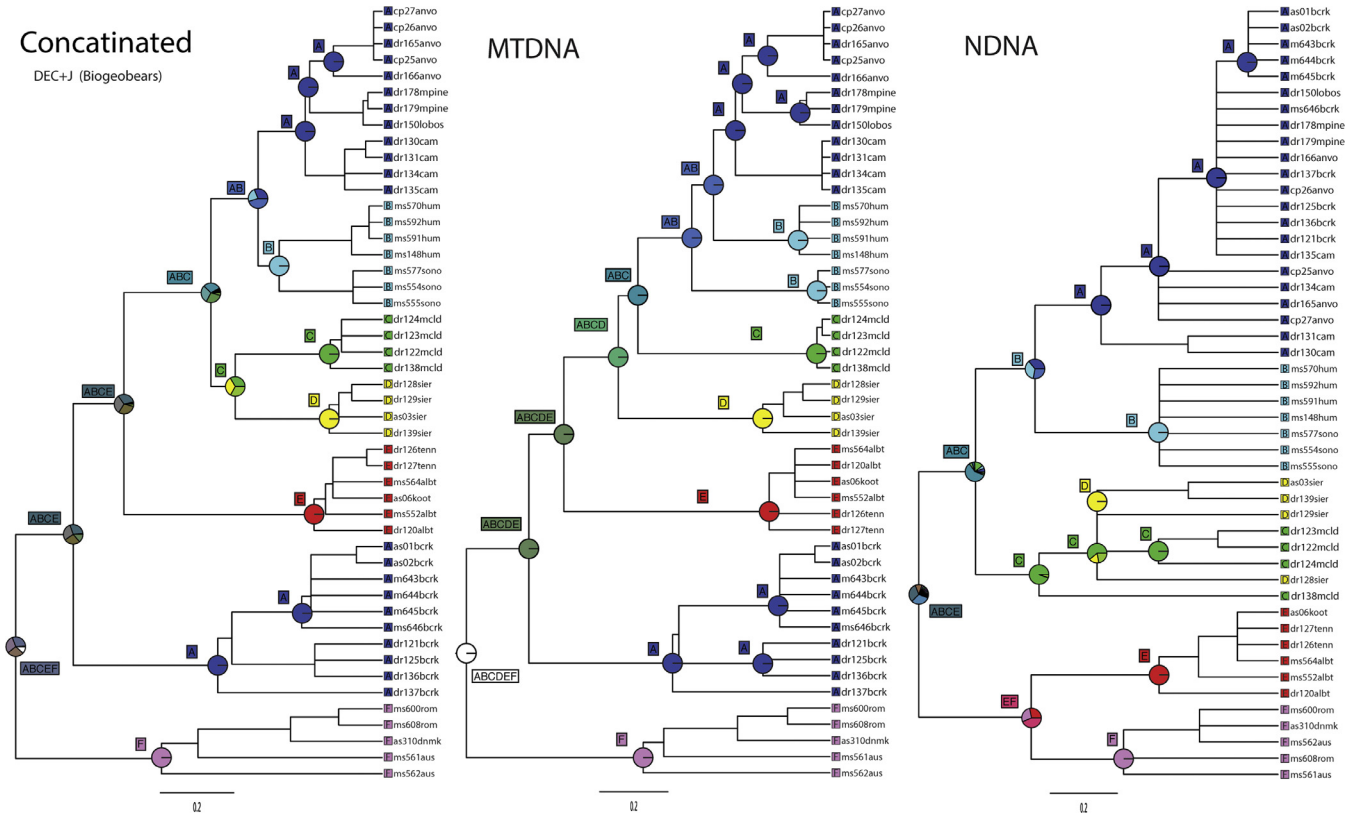

Fig. 9. Maximum likelihood trees, Likelihood analysis of ancestral geographic range using DEC+J model implemented in Biogeobears for all datasets (Concatinated, mtDNA and nDNA). Pie graphs on node indicated relative probability of ancestral geographic range colors and letters correspond to ranges indicated in map (Fig. 1). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2004), but such disagreement can be a fruitful source of information about the process of speciation (Patton and Smith, 1994; Rubinoff and Holland, 2005). Our results clearly demonstrate that both genomes need to be sampled and considered in light of ecological and morphological data before a meaningful designation of taxa can be rendered and a basic understanding of their evolution elucidated. Without all of these data sources, the intriguing and inconsistent patterns of inheritance would not have been evident. The goal of systematics is, ultimately, to understand relationships between taxa and our results clearly show why this would be impossible and the results of either genome misleading, without a combined analysis. The mitochondrial lineages on redwood are phylogenetically very distant from those on Monterey pine.

In the context of their ecology, females of both species must be restricted to their original habitats and hostplants and thus may reflect the original dispersal pattern. This pattern, based on the phylogeny, suggests that the redwood-feeding species is branching basally and that there has been recent contact with the more derived, and recently arriving, Monterey pine species. While the redwood-feeding species' mtDNA suggests an earlier branching origin and long-term isolation from other *Syndemis* species, less discriminate males apparently have caused nDNA to be mixed across these geographically proximate taxa, confounding the original biogeography.

4.1. Biogeography

An examination of the paleoflora bears directly on the patterns of diversification suggested by the mtDNA and supports a much older origin for the sequoia-feeding species. Broadly speaking, Taxodiaceae (including *Sequoia*) is a much older lineage than Pinoideae, dating from Eocene or older in what is now eastern Oregon and Idaho (Axelrod, 1976). *Sequoia* migrated into northern California by the late Miocene/early Pliocene. The original *Syndemis* lineage could have relied on *Sequoia* as early as the Eocene/Oligocene. If this were the case, it would suggest that *Sequoia*'s expansion into the Palearctic occurred later, as there is no fossil record of *Sequoia* in the Old World, though sister taxa persist in China. This supports the idea that *Syndemis* evolved on *Sequoia* and spread to the Palearctic before further diversifying in North America, as suggested by mtDNA.

Regarding the Monterey pine-feeding species, a very different paleobotanic history corroborates the more recent arrival of the moth and its secondary parapatry with the *Sequoia*-feeding species. Closed-cone pine species (Oocarpeae), the precursors of *P. radiata*, are numerous and widely distributed in Mexico and evidently evolved there. Their northward and coastal establishment occurred by the Miocene (Axelrod, 1967, 1980). In California, their occurrence is documented by fossil cones 9 to 3.5 myBP in the Ventura Basin and Tomales Bay- Drakes Bay of Marin Co., California. The northward extension of Oocarpeae was enhanced by the 250–400 km northward movement of the Pacific Coastal Plate during the Miocene (Hornafius et al., 1986). Fossil records for *P. radiata (muricata)*, and its extinct precursors, occur west of the San Andreas fault at localities that were situated far to the south of their modern range during the early Miocene. Our analysis of the non-*Sequoia* feeding *Syndemis* species, based on the phylogeny and hostplants present during these periods, suggests speciation first on conifer and dicot hosts as a generalist, then *Abies* (the host for northern California populations), and finally to *Pinus radiata* with the final expansion back into central California where the relictual *Sequoia*-feeding populations were already restricted.

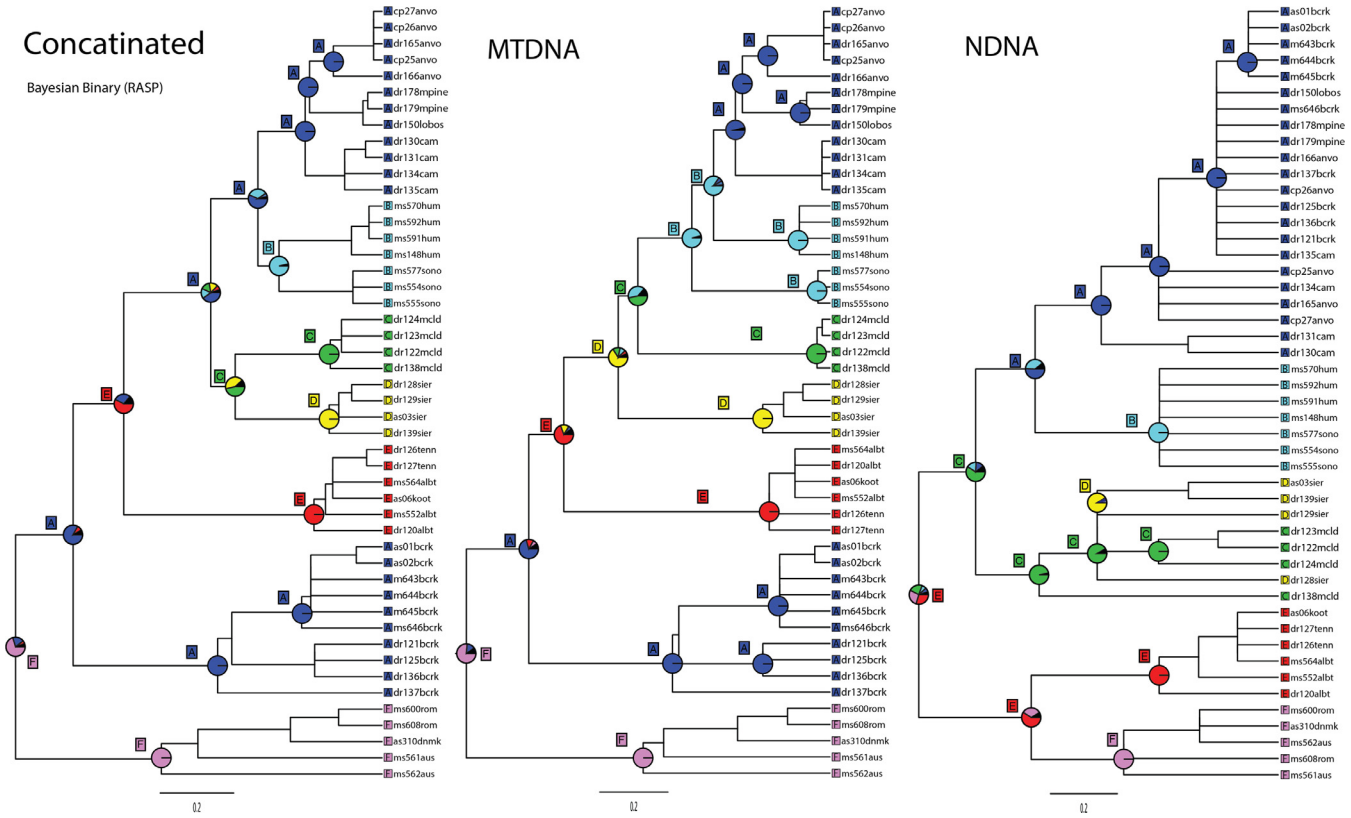

Fig. 10. Maximum likelihood trees, Likelihood analysis of ancestral geographic range using Bayesian Binary model implemented in RASP for all datasets (Concatinated, mtDNA and nDNA). Pie graphs on node indicated relative probability of ancestral geographic range colors and letters correspond to ranges indicated in map (Fig. 1). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5. Conclusion

Based on the earlier branching position of redwood-feeding *Syndemis*' mtDNA, we suggest that this is not a case of incomplete gene sorting, but rather a stable incongruence between the genomes that likely has been maintained since the redwood populations began a secondary exchange of nuclear genetic material with newly-arrived neighboring Monterey pine feeding populations. While the term 'mitochondrial sweep' is frequently used to explain the complete introgression of mtDNA from one species to another, in *Syndemis*, it appears that the nDNA of the redwood-feeders was swept by frequent introgression with neighboring MP feeding populations, and further obscured by a lack of overall differentiation in nDNA across the genus. Alternatively, very strong philopatric female behavior and hostplant-based selection allowed mtDNA to remain isolated; thus the two genomes effectively operate independently from an evolutionary perspective. This cryptic pattern of divergence and reticulation, and the implications it has for the evolution of species, and our ability to accurately discern their true origins, bears further study.

Funding

This project was supported by a U.S. National Science Foundation PEET grant DEB-9521835 to Felix Sperling and Jerry Powell, with additional support from the Monterey Pine Forest Ecology Cooperative and a U.S. National Science Foundation grant DEB-0918341 to D. Rubinoff. Partial funding was provided by the College of Tropical Agriculture and Human Resources, University of Hawaii at Manoa, and USDA Cooperative State Research, Education and Extension (CSREES) project HAW00942-H administered by the College of Tropical Agriculture and Human Resources, University of Hawaii.

Acknowledgments

Ron Wielgus made extensive efforts to find and collect a new species in Humboldt County, CA and provided samples essential to this research. We are sincerely grateful to the following individuals for their assistance in providing specimens for this project: Greg Pohl (Forest Service Canada), Peter Huemer (Tiroler Landesmuseum in Innsbruck, Ole Karsholt (Natural History Museum of Denmark), Jason Dombroskie (Cornell University), Felix Sperling (The University of Alberta). G. E. Daterman (Pacific Northwest Forest and Range Experiment Station, Corvallis Oregon) provided lures essential to collecting most populations of *Syndemis*. We thank Art Fong and the California State Park system for permission to collect in Ano Nuevo and Pt. Lobos State parks.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2017.01.010.

References

Axelrod, D.I., 1976. History of Coniferous Forests. University of California Publlications in Botany, California and Nevada.
Axelrod, D.I., 1967. Evolution of the Californian closed-cone pine forest. In: Philbrick, R.N. (Ed.), Proceedings of the Symposium on the Biology of the California Islands, pp. 93–149.
Axelrod, D.I., 1980. History of the maritime closed-cone pines, Alta and Baja California, vol. 120. Univ of California Press.

Ballard, J.W.O., Whitlock, M.C., 2004. The incomplete natural history of mitochondria. Mol. Ecol. 13, 729–744.

Bouckaert, R.R., 2010. DensiTree: making sense of sets of phylogenetic trees. Bioinformatics 26, 1372–1373.

Bradley, J.D., Temewan, W.G., Smith, A., 1973. British Tortricoid Moths, Cochylidae and Tortricidae:: Tortricinae. Johnson Rerint Co. & E. W. Claasey Ltd., London. viii + 251 pp.

Breton, S., Milani, L., Ghiselli, F., Guerra, D., Stewart, D.T., Passamonti, M., 2014. A resourceful genome: updating the functional repertoire and evolutionary role of animal mitochondrial DNAs. Trends Genet. 30 (12), 555–564.

Brown, J.W., 2005. Tortricidae (Lepidoptera). In: World Catalog of Insects 5: 1-741. Apollo Books Aps, Stenstrup, Denmark.

Brown, J.W., Powell, J.A., 2000. Systematics of Anopina Obratzov (Lepidoptera: Tortricidae: Euliini), vol. 120. University of California Press, Monograph, 2000. 128pp.

Burgess, T., Wingfield, M.J., 2002. Quarantine is important in restricting the spread of exotic seed-borne tree pathogens in the southern hemisphere. Int. For. Rev. 4, 56–65.

Chapman, P.J., Lienk, S.E., 1971. Tortricid Fauna of Apple in New York. N. Y. State Agric. Exp. Sta., Cornell U., Geneva, N.Y.

Cho, S.W., Mitchell, A., Regier, J.C., Mitter, C., Poole, R.W., Friedlander, T.P., Zhao, S. W., 1995. A highly conserved nuclear gene for low-level phylogenetics – elongation Factor-1-Alpha recovers morphology-based tree for heliothine moths. Mol. Biol. Evol. 12, 650–656.

Clement, M., Snell, Q., Walke, P., Posada, D., Crandall, K., 2002. TCS: estimating gene genealogies. In: Proc 16th Int Parallel Distrib Process Symp 2, p. 184.

Cognato, A.I., 2006. Standard percent DNA sequence difference for insects does not predict species boundaries. J. Econ. Entomol. 99, 1037–1045.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. JModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9. 772-772.

Daterman, G.E., Sower, L.L., Stevens, R.E., Fellin, D.G., 1995. Pheromone chemistry and response behavior of conifer-feeding Choristoneura in the western United States. pp. 69–83. In: Powell, J.A. (Ed.), Biosystematic Studies of Conifer-feeding Choristoneura (Lepidoptera: Tortricidae) in the Western United States., vol. 115. U. Calif. Publ. Entomol., p. 275 pp.

Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24 (6), 332–340.

Dombroskie, J.J., 2011. Aspects of Archipine Evolution (Lebidoptera:Tortricidae) Ph. D. disertation. Univeristy of Alberta.

Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5, 1–19.

Farrell, B.D., 1998. "Inordinate Fondness" explained: Why are there so many beetles? Science 281, 555–559.

Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Biol. Biotech. 3, 294–299.

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

Hendrix, R., Fleck, J., Schneider, W., Schneider, C., Geller, D., Avci, A., Olgun, K., Steinfartz, S., 2014. First comprehensive insights into nuclear and mitochondrial DNA based population structure of Near East mountain brook newts (Salamandridae: genus Neurergus) suggest the resurrection of *Neurergus derjugini*. Amphibia-Reptilia 35 (2), 173–187.

Hillis, D.M., Dixon, M.T., 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. Q. Rev. Biol. 66, 411–453.

Horak, M., 1999. The Tortricoidea: 199–215. In: Kristensen, N.P. (Ed.), Lepidoptera, Moths and Butterflies, Volume 1: Evolution, Systematics, and Biogeography, W.

Hornafius, J.S., Luyendyk, B.P., Terres, R.R., Kammerling, M.J., 1986. Timing and extent of Neogene tectonic rotation in western Transverse Ranges, California. Geol. Soc. Am. Bull. 97, 1476–1487.

Hurst, G.D., Jiggins, F.M., 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. P. Roy. Soc. B 272, 1525–1534.

Janz, N., 2011. Ehrlich and Raven revisited: mechanisms underlying codiversification of plants and enemies. Annu. Rev. Ecol. Evol. Syst. 42 (1), 71.

Kodandaramaiah, U., Simonsen, T.J., Bromilow, S., Wahlberg, N., Sperling, F., 2013. Deceptive single-locus taxonomy and phylogeography: Wolbachia-associated divergence in mitochondrial DNA is not reflected in morphology and nuclear markers in a butterfly species. Ecol. Evol. 3 (16), 5167–5176.

Lanfear, R., Calcott, B., Ho, S.Y., Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29 (6), 1695–1701.

Lanfear, R., Calcott, B., Kainer, D., Mayer, C., Stamatakis, A., 2014. Selecting optimal partitioning schemes for phylogenomic datasets. BMC Evol. Biol. 14 (1), 82.

Leaché, A.D., 2010. Species trees for spiny lizards (genus *Sceloporus*): identifying points of concordance and conflict between nuclear and mitochondrial data. Mol. Phylogenet. Evol. 54 (1), 162–171.

Leigh, J.W., Bryant, D., 2015. PopART: full-feature software for haplotype network construction. Methods Ecol. Evol. 6, 1110–1116.

Matzke, N.J., 2013. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. Front. Biogeogr. 5 (4), 242–248.

Matzke, N.J., 2014. Model selection in historical biogeography reveals that founderevent speciation is a crucial process in island clades. Syst. Biol., syu056

Moulton, J.K., Wiegmann, B.M., 2004. Evolution and phylogenetic utility of CAD (rudimentary) among Mesozoic-aged Eremoneuran Diptera (Insecta). Mol. Phylogenet. Evol. 31, 363–378.

Ohshima, I., Yoshizawa, K., 2010. Differential introgression causes genealogical discordance in host rases of *Acrocercops transecta* (Insecta: Lepidoptera). Mol. Ecol. 19, 2106–2119.

Patton, J.L., Smith, M.F., 1994. Paraphyly, polyphyly, and the nature of species boundaries in pocket gophers (genus Thomomys). Syst. Biol. 43, 11–26.

Peters, J.L., Bolender, K.A., Pearce, J.M., 2012. Behavioural vs. molecular sources of conflict between nuclear and mitochondrial DNA: the role of male-biased dispersal in a Holarctic sea duck. Mol. Ecol. 21, 3562–3575. http://dx.doi.org/10.1111/j.1365-294X.2012.05612.x.

Powell, J.A., De Bendicts, J.A., 1995. Biological relationships. Host tree preferences and isolation by pheromones among allopatric and sympatric populations of western *Choristoneura*. In: Powell, J.A. (Ed.), Biosystematic Studies of Conifer-Feeding *Choristoneura* (Lepidoptera: Tortricidae) in the Western United States. University of California Press, Berkeley California, pp. 21–68.

Paradis, E., Claude, J., Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290.

Prentice, R.M., 1966. Forest Lepidoptera of Canada Recorded by the Forest Insect Servey. Canada Department of Forestry, Ottawa.

Rambaut, A., 2009. FigTree, Version 1.3.1. University of Edenburg.

Rambaut, A., Suchard, M.A., Xie, D., Drummond, A.J., 2014. Tracer v1.6. Available from: <http://beast.bio.ed.ac.uk/Tracer>.

Ree, R.H., Moore, B.R., Webb, C.O., Donoghue, M.J., 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. Evol. Biol. 59, 2299–2311.

Ree, R.H., Smith, S.A., 2008. Maximum-likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Syst. Biol. 57, 4–414.

Roe, A., Sperling, F.A.H., 2007. Mol. Ecol.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.

Rubinoff, D., 2006. Utility of mitochondrial DNA barcodes in species conservation. Conserv. Biol. 20, 1026–1033.

Rubinoff, D., Cameron, S., Will, K., 2006. A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. J. Hered. 97, 581–594.

Rubinoff, D., Holland, B., 2005. Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. Syst. Biol. 54, 952–961.

Rubinoff, D., Sperling, F.A.H., 2002. Evolution of ecological traits and wing morphology in *Hemileuca* (Saturniidae) based on a two gene phylogeny. Mol. Phylogenet. Evol. 25, 70–86.

Schliep, K.P., 2011. Phangorn: phylogenetic analysis in R. Bioinformatics 27 (4), 592–593.

Shaw, K.L., 2002. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. Proc. Natl. Acad. Sci. 99 (25), 16122–16127.

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

Simon, C., Buckley, T.R., Frati, F., Stewart, J., Beckenbach, A., 2006. Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. Annu. Rev. Ecol. Syst. 37, 545–579. + 45 pages of on-line supplement.

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.

Slatkin, M., 1985. Gene flow in natural populations. Annu. Rev. Ecol. Syst. 16, 393–430.

Stamatakis, A., Hoover, P., Rougemont, J., 2008. A fast bootstrapping algorithm for the RAxML web-servers. Syst. Biol. 57 (5), 758–771.

Sukumaran, J., Holder, M.T., 2010. DendroPy: a Python library for phylogenetic computing. Bioinformatics 26, 1569–1571.

Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Tosi, A.J., Morales, J.C., Melnick, D.J., 2000. Comparison of Y chromosome and mtDNA phylogenies lead to unique inferences of macaque evolutionary history. Mol. Phylogent. Evol. 17, 133–144.

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.

Wahlberg, N., Weingartner, E., Warren, A.D., Nylin, S., 2009. Timing major conflict between mitochondrial and nuclear genes in species relationships of Polygonia butterflies (Nymphalidae: Nymphalini). BMC Evol. Biol. 9 (1), 1.

Wilgenbusch, J.C., Warren, D.L., Swofford, D.L., 2004. AWTY: A system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference. <http://ceb.csit.fsu.edu/awty>.

Will, K.W., Mishler, B.D., Wheeler, Q.D., 2005. The perils of DNA barcoding and the need for integrative taxonomy. Syst. Biol. 54, 844–851.

Yu, Y., Harris, A.J., He, X.-J. 2011. RASP (Reconstruct Ancestral State in Phylogenies) 2.1b. <http://mnh.scu.edu.cn/soft/blog/RASP>.

Zwickl, D.J., 2006. Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets under the Maximum Likelihood Criterion Ph. D. dissertation. The University of Texas at Austin.