Species tree estimation of North American chorus frogs (Hylidae: *Pseudacris*) with parallel tagged amplicon sequencing

CrossMark

Lisa N. Barrow [*], Hannah F. Ralicki, Sandra A. Emme, Emily Moriarty Lemmon

Department of Biological Science, Florida State University, 319 Stadium Drive, P.O. Box 3064340, Tallahassee, FL 32306-4340, USA

A B S T R A C T

The field of phylogenetics is changing rapidly with the application of high-throughput sequencing to non-model organisms. Cost-effective use of this technology for phylogenetic studies, which often include a relatively small portion of the genome but several taxa, requires strategies for genome partitioning and sequencing multiple individuals in parallel. In this study we estimated a multilocus phylogeny for the North American chorus frog genus *Pseudacris* using anonymous nuclear loci that were recently developed using a reduced representation library approach. We sequenced 27 nuclear loci and three mitochondrial loci for 44 individuals on 1/3 of an Illumina MiSeq run, obtaining 96.5% of the targeted amplicons at less than 20% of the cost of traditional Sanger sequencing. We found heterogeneity among gene trees, although four major clades (Trilling Frog, Fat Frog, *crucifer*, and West Coast) were consistently supported, and we resolved the relationships among these clades for the first time with strong support. We also found discordance between the mitochondrial and nuclear datasets that we attribute to mitochondrial introgression and a possible selective sweep. Bayesian concordance analysis in BUCKy and species tree analysis in *BEAST produced largely similar topologies, although we identify taxa that require additional investigation in order to clarify taxonomic and geographic range boundaries. Overall, we demonstrate the utility of a reduced representation library approach for marker development and parallel tagged sequencing on an Illumina MiSeq for phylogenetic studies of non-model organisms.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

The fields of phylogenetics and phylogeography are undergoing a transformation driven by the amount of data being collected with high-throughput sequencing approaches (reviewed in Lemmon and Lemmon (2013) and McCormack et al. (2013a)). Until recently, studies have often been limited to a single or few genetic loci that are commonly used and for which primers have already been developed. It is well known, however, that discordance among gene trees due to incomplete lineage sorting, hybridization, or gene duplication make it necessary to incorporate multilocus nuclear data in order to recover the true species history (Brito and Edwards, 2009; Edwards, 2009; Knowles, 2009; Maddison, 1997; Pamilo and Nei, 1988; Slowinski and Page, 1999). The development of informative genetic markers that are appropriate for the evolutionary scale of each question remains a significant challenge to these fields, although applications of recent sequencing technologies in combination with genome partitioning strategies are allowing

researchers to make rapid progress (Davey et al., 2011; Faircloth et al., 2012; Lemmon et al., 2012; Mamanova et al., 2010).

In addition to identifying appropriate genetic markers, the development of analytical methods that address issues leading to discordance between gene trees and species trees is an area of active research (e.g., Ané et al., 2007; Heled and Drummond, 2010; Kubatko et al., 2009; Liu and Pearl, 2007; Liu et al., 2009). Traditional phylogenetic analyses that incorporate multiple loci have often employed concatenated datasets that assume a single topology underlying the species history. It has been demonstrated, however, that concatenation can produce misleading results particularly when species are undergoing rapid divergence (i.e., when branch lengths are short; Belfiore et al., 2008; Degnan and Rosenberg, 2006; Edwards et al., 2007; Kubatko and Degnan, 2007). As methods continue to develop, it is important to evaluate their performance on both simulated and empirical datasets and compare across different methods to provide confidence in the estimated evolutionary history (Chung and Ané, 2011; Leaché and Rannala, 2011; Lee et al., 2012; Liu et al., 2009; Weisrock et al., 2012).

The North American chorus frog genus *Pseudacris* has been a valuable system for speciation studies, providing evidence for reproductive character displacement (Fouquette, 1975; Lemmon,

* Corresponding author.
E-mail addresses: barrow@bio.fsu.edu (L.N. Barrow), hfr08@my.fsu.edu (H.F. Ralicki), sandysu57@gmail.com (S.A. Emme), chorusfrog@bio.fsu.edu (E.M. Lemmon).

2009), hybridization (Gartside, 1980; Lemmon et al., 2007a), reinforcement (Lemmon, 2009; Lemmon and Lemmon, 2010), and allopatric divergence driven by climate and geology (Lemmon et al., 2007b). Critical to making inferences about the evolution of traits and the timing of divergences is a well-characterized phylogenetic framework. Previously the taxonomy of this group has been the source of some confusion, partially due to conservative morphology and subtle acoustic signal differences (e.g., Neill, 1949; Platz, 1989; Platz and Forester, 1988; Smith and Smith, 1952). The most recent phylogeny for the group was estimated using 2.4 kb of mitochondrial DNA (Lemmon et al., 2007a; Moriarty and Cannatella, 2004). The incorporation of nuclear loci is essential to understand the evolutionary relationships in this group particularly given the apparent evidence for mitochondrial introgression between some species (e.g., *P. clarkii* and *P. maculata*; Lemmon et al., 2007a).

The goal of this study is to estimate a multilocus phylogeny for *Pseudacris* chorus frogs using mitochondrial DNA and newly-developed anonymous nuclear loci in order to: (1) validate a method of marker development for phylogenetic studies of non-model organisms; (2) identify and evaluate discrepancies between mitochondrial and nuclear data; (3) compare traditional phylogenetic analysis methods with recently-developed species tree methods; and (4) determine the evolutionary relationships among species within this genus. This work demonstrates the efficacy of using a reduced representation library (RRL) approach to identify markers for shallow phylogenetic scales as in Lemmon and Lemmon (2012). We also generate a new working phylogeny for *Pseudacris* and demonstrate the utility of parallel tagged sequencing (PTS, O'Neill et al., 2013; also known as targeted amplicon sequencing, Bybee et al., 2011a) on an Illumina MiSeq platform for phylogenetic studies of non-model organisms.

2. Material and methods

2.1. Sampling

We obtained *Pseudacris* tissue samples from across its range from previous field collections and through museum loans (Table 1; Fig. 1). Tissues were either frozen in liquid nitrogen or preserved in tissue buffer (20% DMSO, 0.25 M EDTA, salt-saturated) or 95% ethanol and held at −80 °C at Florida State University (FSU). The 44 samples used in this study consisted of at least two individuals per species of *Pseudacris*, including previously described subspecies and mitochondrial lineages (Lemmon et al., 2007a; Recuero et al., 2006). We refer to the clades described by Moriarty and Cannatella (2004) throughout this study: (1) West Coast Clade (*cadaverina*, *regilla*, *sierra*, and *hypochondriaca*); (2) Fat Frog Clade (*ornata*, *streckeri*, and *illinoensis*); (3) *crucifer* Clade (*ocularis* and *crucifer*); and (4) Trilling Frog Clade (*brachyphona*, *brimleyi*, *clarkii*, *feriarum* (inland and coastal clades), *fouquettei*, *kalmi*, *maculata*, *nigrita*, and *triseriata*). We also include two individuals each of *Acris gryllus* and *Hyla cinerea* as outgroups.

2.2. DNA extraction, amplification, library preparation, and sequencing

We extracted genomic DNA from liver tissue or leg muscle following the protocols of the E.Z.N.A.® Tissue DNA Kit (Omega Bio-Tek, Norcross, GA, USA). The anonymous nuclear loci used for this study were developed using an RRL approach described in detail in Lemmon and Lemmon (2012). Briefly, genomic DNA from two species (*P. feriarum* and *P. maculata*) was digested with restriction enzymes, size selected to a range of ~500–600 bp on an agarose gel, and each species sequenced on an Illumina HiSeq 2000 lane (100 bp paired end). Primers were designed for 187 orthologous loci that were single-copy and exhibited polymorphism among

species. These primers were screened across all species of *Pseudacris* as well as the outgroup taxa, and 27 loci that amplified as single bands in the correct size range were selected for sequencing in this study. We also include three mitochondrial loci amplified using previously designed primers: a 707-bp portion of cytochrome b (cytb), a 710-bp portion of cytochrome oxidase I (COI), and a 627-bp portion of NADH dehydrogenase subunit 2 (ND2). All primer sequences and references are listed in Supplementary Table S1.

For each individual, the 30 loci were amplified in separate PCR reactions that consisted of $1\times$ Go Taq® Reaction Buffer (Promega), 0.08 mM dNTPs, 0.4 U Go Taq® DNA Polymerase, 0.2 μM each primer, and 18 ng template DNA in a total volume of 15 μL. Amplification was performed on a Bio-Rad DNA Engine Tetrad® 2 thermal cycler using the following temperature profiles: for nuclear loci (nDNA), an initial denaturation at 95 °C for 2 min, then 35 cycles of 95 °C for 30 s, annealing at primer specific temperatures (Table S1) for 30 s, and 72 °C for 1 min, then a final extension at 72 °C for 5 min, and then held at 4 °C; for mitochondrial loci (mtDNA), an initial denaturation at 94 °C for 2 min, then 35 cycles of 94 °C for 30 s, 42 °C for 30 s, and 72 °C for 1 min, then a final extension at 72 °C for 4 min, and then held at 4 °C. We verified PCR success by electrophoresing 2 μL PCR product in a 1% $1\times$ TAE agarose gel for 20 min at 120 V and visualizing with an Ultra-Violet Products (UVP, LLC, Upland, CA, USA) transilluminator to identify single bands. Successful PCR products were purified using a magnetic bead wash with 4 μL Agencourt® SprintPrep® beads (Beckman Coulter, Inc.). Cleaned products were quantified with a NanoDrop® ND-1000 Spectrophotometer (Thermo Fisher Scientific, Inc.) and loci were pooled within each individual at equal concentrations to a final volume and concentration of 50 μL at 20 ng/μL (1 μg total DNA to begin library preparation).

For each individual, the pooled PCR products were sheared to a size distribution centered on ~250–300 bp using a Covaris™ E220 focused ultra-sonicator (Covaris, Inc., Woburn, MA, USA). Libraries were prepared for each individual using unique barcode adapters and a protocol modified from Meyer and Kircher (2010). The libraries were quantified using a high sensitivity assay on a Qubit™ fluorometer (Invitrogen, Carlsbad, CA, USA) and then pooled at equimolar concentrations along with libraries from two other projects. The pooled libraries were assayed on an Agilent DNA 1000 bioanalysis chip to determine the final size range and concentration for sequencing. The final pool was diluted to 1 nM and was sequenced in a single Illumina MiSeq run (150 bp paired end) at the FSU Biology Core Facility. To obtain a complete data matrix for at least one mtDNA gene (cytb) for all individuals, cleaned PCR products from 10 individuals with low coverage were Sanger sequenced in both directions on an Applied Biosystems 3730 Genetic Analyzer using Big-Dye v. 3.1 terminator chemistry at the FSU Sequencing Facility.

2.3. Assembly, locus sorting, and alignment

Sequence reads were sorted bioinformatically with a script written in Java by A.R. Lemmon, which sorted reads by individual using the unique barcode adapter sequences. We then assembled the reads for each individual using SeqMan NGEN v. 2.2.0 (DNA-STAR, Inc., Madison, WI) with the following settings to ensure high quality: specific match size set to 21, default quality set to 30, and maximum gap set to 2. Consensus sequences for each contig were aligned to the primer sequences in Sequencher v. 4.7 (Gene Codes Corporation, Ann Arbor, MI) to sort the contigs by locus and construct complete assemblies for each locus. We used a conservative threshold value of $10\times$ coverage and only included loci for each individual that had complete sequences at this threshold to minimize missing data within alignments for

Table 1
List of samples in this study including ID numbers used for analysis, field numbers, museum voucher numbers, and locality information. All taxa are *Pseudacris* except where genera are listed. Museum abbreviations are as follows: Sternberg Museum of Natural History, Fort Hays State University (MHP), and Texas Natural History Collection, University of Texas, Austin (THNC). **I** = Inland clade, **C** = Coastal clade, based on 2.4 kb of mtDNA (Lemmon et al., 2007a). An "n/a" under Museum No. indicates that vouchers and additional specimen information are available from the authors.

ID	Taxon	Field no.	Museum no.	State	County	Latitude	Longitude
1	*brachyphona*	ECM0111	TNHC62305	AL	Elmore	32.5175	−86.0071
2	*brachyphona*	ECM5218	n/a	TN	Campbell	36.4167	−84.2838
3	*brimleyi*	ECM0079	TNHC62329	NC	Pitt	35.7006	−77.4094
4	*brimleyi*	ECM0469	TNHC63573	NC	Craven	35.1892	−77.0814
5	*cadaverina*	ECM0150	TNHC62247	CA	San Bernardino	34.1132	−117.1422
6	*cadaverina*	SSS32487	MHP15045	CA	San Luis Obispo	35.0695	−120.1745
7	*clarkii*	ECM0102	TNHC62253	KS	Chautauqua	37.0044	−96.2764
8	*clarkii*	ECM2478	TNHC65763	TX	Travis	30.0205	−97.6946
9	*crucifer*	Y-1	TNHC62369	FL	Lake	29.0833	−81.5833
10	*crucifer*	ECM3702	n/a	IN	LaGrange	41.6883	−85.3025
11	*feriarum* **I**	ECM0387	TNHC63123	AL	Macon	32.5290	−85.6016
12	*feriarum* **I**	ECM0448	TNHC63362	GA	Heard	33.2765	−85.1211
13	*feriarum* **C**	ECM0464	TNHC63567	NC	Davie	35.8982	−80.5764
14	*feriarum* **C**	ECM1011	TNHC63645	VA	Prince Edward	37.0973	−78.4770
15	*kalmi*	ECM0162	TNHC62354	MD	Kent	39.3122	−75.8485
16	*kalmi*	ECM1080	TNHC63135	DE	Sussex	38.7459	−75.3809
17	*maculata*	ECM0099	TNHC62377	KS	Douglas	38.9501	−95.2871
18	*maculata*	ECM2746	n/a	IL	Randolph	38.0442	−90.0439
19	*nigrita*	ECM0024	TNHC62364	FL	Brevard	28.2006	−80.8678
20	*nigrita*	ECM0372	TNHC63201	FL	Jefferson	30.1981	−84.0500
21	*ocularis*	ECM0045	TNHC62234	SC	Barnwell	33.1606	−81.6908
22	*ocularis*	ECM5886	n/a	FL	Nassau	30.5136	−81.8748
23	*regilla*[a]	JTC3276	MHP14888	OR	Clackamas	45.4299	−122.3736
24	*regilla*[a]	JTC3382	n/a	OR	Benton	44.5350	−123.2572
25	*hypochondriaca*[a]	ECM0147	TNHC62195	CA	San Bernardino	34.1132	−117.1422
26	*hypochondriaca*[a]	SSS32542	MHP15076	CA	Santa Barbara	34.7878	−120.4530
27	*sierra*[a]	JTC3238	MHP14886	CA	San Luis Obispo	35.3922	−120.4736
28	*sierra*[a]	JTC3409	MHP15178	CA	Modoc	41.4643	−120.5191
29	*sierra*[a]	ECM2697	n/a	CA	Sacramento	38.5327	−121.2574
30	*sierra*[a]	ECM2694	n/a	CA	Mendocino	39.7011	−123.6595
31	*fouquettei*	ECM0011	TNHC62255	AR	Craighead	35.8546	−90.6626
32	*fouquettei*	ECM0029	TNHC62265	LA	East Baton Rouge	30.6889	−90.8894
33	*triseriata*	ECM5403	n/a	IN	Posey	38.0702	−87.9573
34	*triseriata*	ECM3468	n/a	IN	Jasper	41.1561	−86.9203
35	*Acris gryllus*	ECM5835	n/a	FL	Liberty	30.0335	−84.9989
36	*Acris gryllus*	ECM5938	n/a	FL	Walton	30.6971	−85.9122
37	*Hyla cinerea*	LNB00326	n/a	FL	Franklin	29.7668	−84.8791
38	*Hyla cinerea*	ECM3184	n/a	IL	Jackson	37.7725	−89.3858
39	*illinoensis*	ECM0001	TNHC62351	AR	Clay	36.3308	−90.1090
40	*illinoensis*	ECM4370	n/a	MO	Mississippi	36.7851	−89.3142
41	*ornata*	ECM0033	TNHC62178	AL	Barbour	32.0369	−85.0889
42	*ornata*	ECM0057	TNHC62185	SC	Barnwell	33.3178	−81.4769
43	*streckeri*	P-3	TNHC62318	TX	Travis	30.3218	−97.8034
44	*streckeri*	MHP8220	MHP8220	KS	Pratt	37.4999	−98.5400

[a] Taxa are labeled based on ranges depicted in Recuero et al. (2006).

each locus. Single nucleotide polymorphisms (SNPs) were identified in SeqMan v. 8.1.3 (DNASTAR, Inc., Madison, WI) using a threshold of 10× coverage and a minimum SNP Filter of 20%. Alleles were phased to haplotype by visually examining each assembly in SeqMan and using the paired end read data to determine which bases were associated together in each allele. When the state of a SNP could not be determined due to low coverage, it was coded as a degenerate base.

Alignments including all individuals (two allele sequences for heterozygotes) for each locus were assembled using MUSCLE v. 3.8.31 with default settings (Edgar, 2004). Alignments were visually inspected in SeaView v. 4.4.0 (Gouy et al., 2010) and primer regions were excluded. Highly variable regions that could not be aligned unambiguously were excluded using Gblocks v. 0.91b (Castresana, 2000) with settings for a less stringent selection: allowing for smaller final blocks, allowing gap positions within the final blocks, and allowing less strict flanking positions. These relaxed settings allow a greater number of contiguous nonconserved positions, allow gaps if they are present in less than half of the sequences, and decrease the required number of sequences with conserved flanking positions. Relaxed block selection has

been shown to perform better for short alignments (400–800 amino acid positions; Talavera and Castresana, 2007).

2.4. Model testing and gene tree analysis

We selected appropriate models of nucleotide evolution for each locus using hierarchical likelihood ratio tests in MrModeltest2 v. 2.3 (Nylander, 2004). Two different nDNA datasets were used in phylogenetic analyses: a multiple-allele dataset including both alleles for heterozygotes, and a single-allele dataset including a randomly-chosen allele for each individual. The single-allele datasets were generated using a script written in Python v. 3.1.2, which randomly selected an allele for each heterozygote and produced datasets for single gene tree, concatenated, and species tree analyses. We estimated gene trees for each locus (three mtDNA and 27 nDNA) using MrBayes v. 3.1.2 (Ronquist and Huelsenbeck, 2003) with locus-specific models (Table 2). We conducted the mtDNA gene tree analyses in two ways: (1) without partitioning, and (2) with the three codon positions partitioned separately. Each analysis consisted of four replicate runs with two Markov chain Monte Carlo (MCMC) chains. Analyses were run for 20 million generations
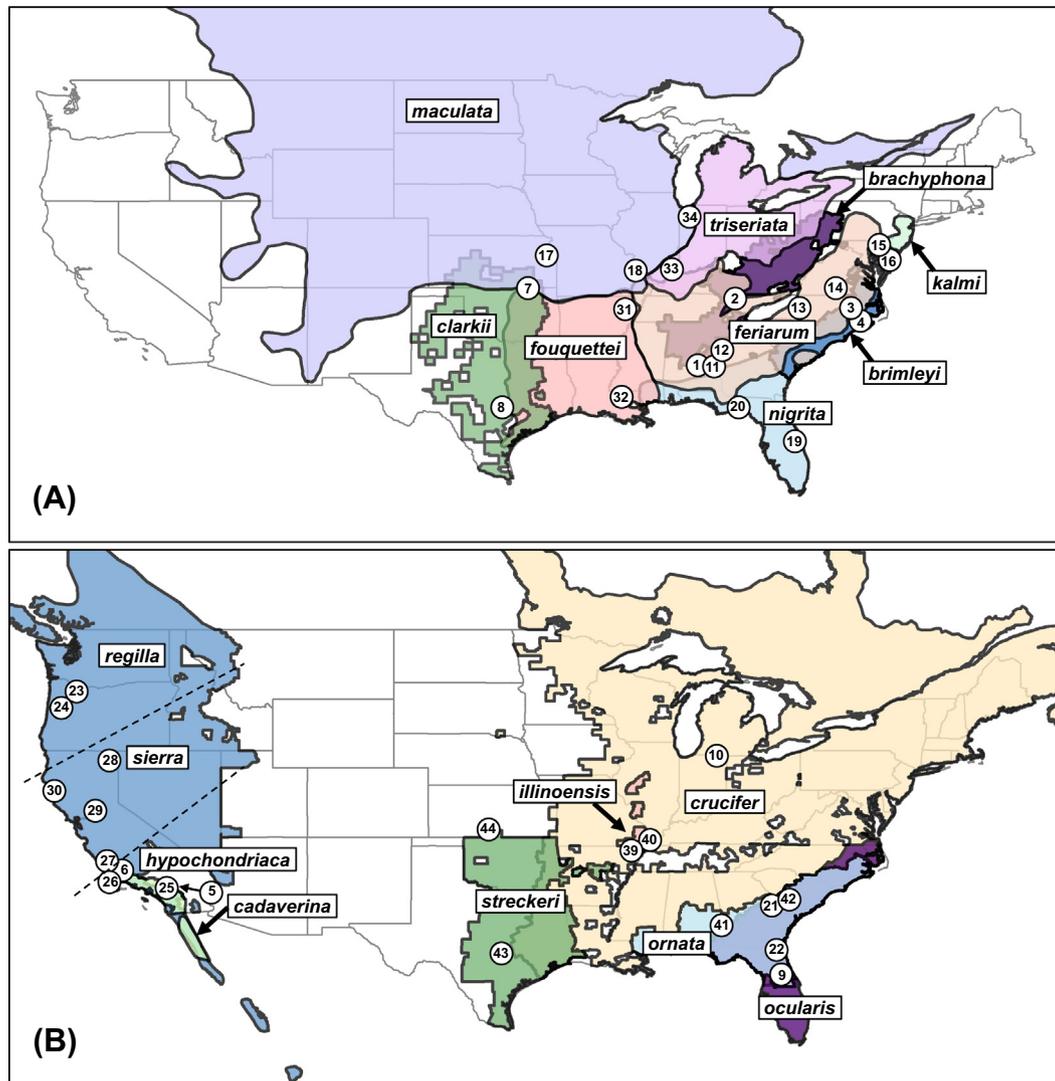
Fig. 1. Range maps of *Pseudacris* species in the U.S. and sample localities in this study. (A) Trilling Frog Clade. (B) West Coast, *crucifer*, and Fat Frog Clades. Circled numbers are sample localities with identifiers (IDs in Table 1) for each individual, which correspond to the taxon numbers in Figs. 2–5. Ranges are based on spatial data from the IUCN Red List (IUCN, 2012) and dashed lines are approximate boundaries redrawn from Recuero et al. (2006).

and sampled every 2000 generations producing a total of 10,000 samples for each of the four runs. Gene tree analyses were performed using the shared High-Performance Computing resources at FSU. We assessed convergence by examining the average standard deviation of split frequencies among runs (all were less than 0.01 except for three of the multiple-allele datasets: Locus 08 = 0.0108, Locus 168 = 0.0178, and Locus 172 = 0.0112), and by visualizing the sampled parameter values in Tracer v. 1.5 (Rambaut and Drummond, 2007). In all analyses, after 10% of the samples were discarded as burn-in, stationarity was reached and the effective sample sizes (ESS) were greater than 200, indicating a sufficient number of independent samples. The remaining 9001 samples from each run (36,004 samples per locus) were summarized to produce a consensus tree with clade credibility values (posterior probabilities).

2.5. Concatenated analysis

We conducted maximum likelihood analyses on separate concatenated datasets for mtDNA and nDNA using RAxML v. 7.4.2 (Stamatakis, 2006). One of the nDNA loci was identified as a potential paralog (see Locus 01 in Table 2) because of its highly variable

alignment in which three *Pseudacris* individuals exhibited high sequence similarity with the outgroup taxa, a pattern not predicted for orthologous loci. This locus was not included in the concatenated or species tree analyses, resulting in a final nDNA dataset of 26 loci. Only the single-allele datasets for each nDNA locus were used for concatenated analysis to avoid making assumptions about which allele from a locus corresponded to alleles at other loci. We executed the rapid hill-climbing algorithm in RAxML, partitioned each dataset by locus, and used the GTR + G model for each partition, which is the most appropriate model available in RAxML given our dataset. We also analyzed the mtDNA dataset with separate partitions for each codon position. We then conducted non-parametric bootstrapping with 1000 replicates and present the support values on the best-scoring maximum likelihood tree.

2.6. Species tree analysis

Species tree analyses were performed using two methods that differ in their assumptions about the cause of gene tree heterogeneity. We first conducted Bayesian concordance analysis in BUCKy v. 1.4.0 (Ané et al., 2007; Larget et al., 2010), which does not make any assumption about the causes of gene tree discordance, but

Table 2
Description of the 27 anonymous nuclear and three mtDNA loci in this study. Product lengths are listed as ranges when there is variation among individuals because of insertions and deletions. The length analyzed is the alignment length after excluding primer regions and highly variable regions using Gblocks (Castresana, 2000). The ingroup-variable and ingroup-informative sites were determined in PAUP* v. 4b10 (Swofford, 2003). Models of nucleotide evolution were chosen using MrModeltest2 (Nylander, 2004).

Locus name	Product length in bp	Length analyzed in bp	No. ingroup informative sites (%)	No. ingroup variable sites (%)	Model of nucleotide evolution
1[a]	552–561	506	128 (25.3)	150 (29.6)	GTR + G
7	504–509	463	44 (9.5)	70 (15.1)	K80 + G
8	558–584	511	37 (7.2)	64 (12.5)	HKY + G
17	586–625	569	37 (6.5)	56 (9.8)	HKY + G
57	434–487	430	81 (18.8)	102 (23.7)	HKY + G
58	561–578	527	68 (12.9)	91 (17.3)	HKY + G
71	461–470	422	42 (9.95)	57 (13.5)	GTR + G
80	584–604	558	58 (10.4)	69 (12.4)	HKY + G
82	564–582	519	29 (5.59)	39 (7.51)	HKY + G
92	547–597	508	35 (6.89)	57 (11.2)	HKY + G
102	583–599	538	57 (10.6)	77 (14.3)	HKY + G
104	581–603	536	51 (9.51)	72 (13.4)	HKY + G
108	598–610	566	58 (10.3)	77 (13.6)	HKY + G
116	578–608	526	67 (12.7)	87 (16.5)	GTR + G
125	472–481	430	32 (7.44)	46 (10.7)	F81 + G
129	518–533	483	47 (9.73)	72 (14.9)	HKY + G
137	558	512	54 (10.6)	73 (14.3)	HKY + G
140	508–553	462	54 (11.7)	76 (16.5)	HKY + G
148	573–586	526	48 (9.13)	67 (12.7)	K80 + G
162	429–439	390	54 (13.9)	82 (21.0)	HKY + G
166	462	417	53 (12.7)	75 (18.0)	GTR + G
167	556–567	507	27 (5.33)	40 (7.89)	HKY + G
168	485–502	459	77 (16.8)	120 (26.1)	K80 + G
169	513–519	463	40 (8.64)	64 (13.8)	HKY + G
172	441–449	402	46 (11.4)	61 (15.2)	SYM + I + G
182	519–524	473	45 (9.51)	64 (13.5)	GTR + G
183	528	482	44 (9.13)	60 (12.5)	HKY + G
COI	710	658	230 (35.0)	238 (36.2)	GTR + I + G
cytb	707	648	228 (35.2)	245 (37.8)	GTR + I + G
ND2	627	580	198 (34.1)	214 (36.9)	GTR + G
TOTAL		15,071	2069 (13.7%)	2665 (17.7%)	

[a] Locus identified as a potential paralog and removed from further phylogenetic analyses.

rather summarizes the posterior distributions of the individual gene trees in order to estimate a primary concordance tree. Each clade is described by a concordance factor (CF), which is defined as the proportion of sampled loci that contain that clade. Strictly speaking, this primary concordance tree is not a species tree, but rather a "dominant tree," depicting the clades whose CFs exceed those of any contradictory clades (Ané et al., 2007). Bayesian concordance analysis requires that the same tips be present in the posterior distributions of trees for every locus, so if an individual was missing data for a given locus, we added a line of question marks to the MrBayes locus data file for that individual. It is expected that the phylogenetic placement of individuals with missing data should be random across the posterior distribution of trees so results will not be affected. We summarized the gene tree distributions obtained for each nDNA locus from the single-allele MrBayes analyses after discarding the 10% burn-in (36,004 trees summarized per locus). We then ran four independent analyses in BUCKy with four MCMC chains for 1 million generations following a burn-in of 100,000 generations. We tested different values for the *a priori* level of discordance among loci as in Weisrock et al. (2012), α = 0.1, 1, 10, and 100 (α = 0 corresponds to the assumption that all loci have the same topology; α = ∞ corresponds to the assumption that gene trees are completely independent). To ensure that missing data did not affect our results, we repeated the BUCKy analyses with complete datasets created by sub-sampling in two ways: the first sub-sampled dataset included 41 individuals (all *Pseudacris* individuals but only one outgroup) and 18 loci, and the second included 20 individuals (one individual per species and one outgroup) with 25 loci.

We also conducted species tree analyses using *BEAST v. 2.0 (Heled and Drummond, 2010), which uses multilocus data to simultaneously co-estimate gene trees embedded within a species tree under a coalescent model. This framework involves making the assumption that all gene tree discordance is due to incomplete lineage sorting. In *BEAST, the term "species" refers to any group of individuals that does not experience gene flow with individuals outside of that group (also described as a population). We conducted analyses with two different types of taxon sets. First, we created taxon sets in which every putative species, subspecies, or mitochondrial lineage (listed in Table 1 as "Taxon") was considered a population. We conducted these "species" tree analyses with both the single-allele and multiple-allele datasets for the 26 nDNA loci such that every population was represented by either 2–4 alleles at every locus (one per individual, single-allele) or 4–8 alleles at every locus (two per individual, multiple-allele). Second, because our sampled individuals were geographically disparate and we were interested in investigating the status of previously described subspecies and mtDNA lineages, we created taxon sets in which every individual was considered a population. We used multiple-allele datasets for the 26 nDNA loci such that every population (i.e., individual) was represented by two alleles at every locus. In this study, we were interested primarily in tree topology and reserve divergence time and population size estimates for future work. We used locus-specific models of nucleotide evolution, choosing either HKY + G or GTR + G with four gamma categories (Table 2). We tested both strict and relaxed lognormal clock models in preliminary analyses and found consistent species tree topologies. For our final analysis, we chose the strict clock model for all loci, a constant species tree population size, and a Yule model species tree prior. We ran the MCMC analysis for one billion generations sampling every 50,000 generations. We assessed convergence by visualizing the sampled parameter values in Tracer
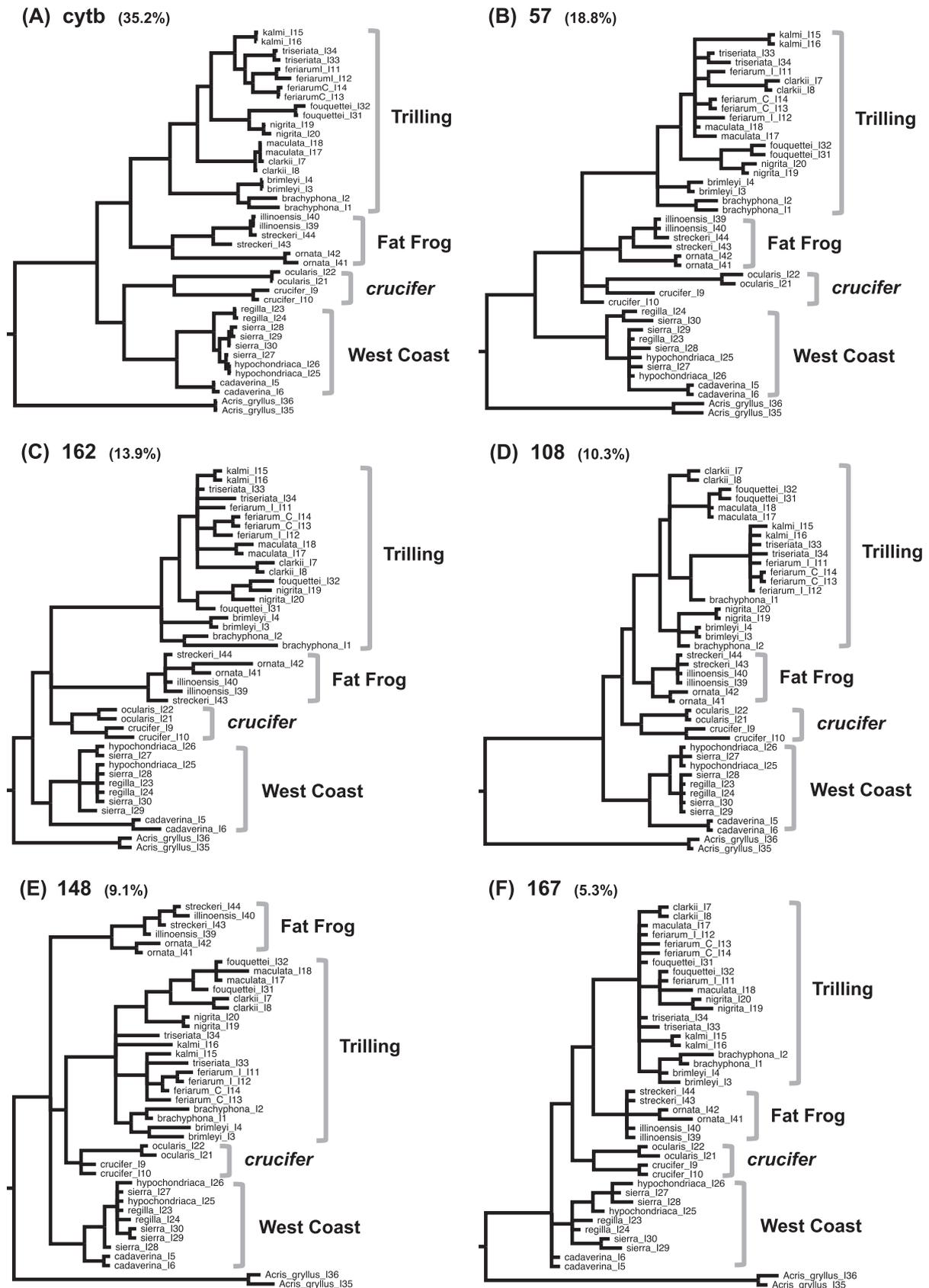
v. 1.5 (Rambaut and Drummond, 2007), and used TreeAnnotator v. 1.6.1 (Drummond and Rambaut, 2007) to visualize the maximum clade credibility species tree after discarding the first 10% of samples as burn-in.

3. Results

3.1. Sequence data obtained

We obtained a total of 5,198,644 reads from the Illumina MiSeq run, which combined individuals from this study with two other projects. We sorted 5,041,216 (97.0%) of the reads by individual barcode and removed the unsorted reads from further analysis. The 44 individuals for this study included a total of 2,254,324 reads with an average of 51,235 ± 17,706 (s.d.) reads per individual. After applying strict settings to ensure high quality, we assembled 605,185 (26.7%) of the reads with an average of 13,754 ± 4954 (s.d.) reads per individual. Despite the relatively low number of assembled reads, we wanted to ensure the accuracy of our alignments and downstream phylogenetic analyses, and we still obtained complete sequences at a threshold of 10× coverage for 1257 (96.5%) of the pooled amplicons (Fig. 2). For the 27 nDNA loci, only a single PCR failed (out of 44 × 27 = 1188 total reactions). When both PCR failures and low coverage sequencing are taken into account, the full dataset of three mtDNA and 27 nDNA loci for 44 individuals was 96.0% complete (Fig. 2). We obtained a complete data matrix for 18 nDNA loci for the 40 ingroup individuals. Sequence data generated by the Illumina MiSeq run, files from the analyses including alignments and trees, and custom scripts used in this study have been deposited in the Dryad data repository (doi:10.5061/dryad.23rc0). Sequences generated in this study have also been deposited on GenBank (accession numbers KJ536151–KJ538148).

3.2. Properties of the loci

Alignments totaling 15,071 bp were analyzed in this study, which included 2665 (17.7%) variable sites and 2069 (13.7%) parsimony-informative sites within the ingroup (Table 2). The three mtDNA loci consisted of 1886 total bp with an average length of 629 bp (range: 580–658). The mtDNA loci contained 697 (37.0%) variable sites and 656 (34.8%) parsimony-informative sites. On average, each mtDNA locus contained 232 variable sites (range: 214–245) and 219 parsimony-informative sites (range: 198–230). The 26 nDNA loci used for the concatenated and species tree analyses consisted of 12,679 total bp with an average length of 488 bp (range: 390–569). The nDNA loci contained 1818 (14.3%) variable sites and 1285 (10.1%) parsimony-informative sites. On average, each nDNA locus contained 70 variable sites (range: 39–120) and 49 parsimony-informative sites (range: 27–81).

3.3. Individual gene trees

The gene trees estimated in MrBayes demonstrate a high level of heterogeneity both in tree topology and branch lengths (Fig. 3). In general, the individual nDNA trees contain less variation and a lack of resolution compared to the mtDNA cytb tree (Fig. 3A). The multiple-allele gene trees include several instances in which

Fig. 2. Locus by individual matrix of sequence data obtained in this study.

Fig. 3. Example gene trees estimated in MrBayes using the single-allele datasets for (A) the mtDNA cytb gene and (B–F) five representative nDNA loci. The locus name is listed with the percentage of ingroup parsimony-informative sites in parentheses.

the two alleles from a single individual are not monophyletic (not shown), suggesting the presence of incomplete lineage sorting and/ or hybridization. Similarly, the single-allele trees have some cases in which species do not exhibit reciprocal monophyly (e.g.,

brachyphona in Fig. 3D). Despite the variation in topology among loci and lower resolution in the nDNA trees, the four major clades within *Pseudacris* described by Moriarty and Cannatella (2004) are consistently supported: the Trilling Frog, Fat Frog, *crucifer*, and West Coast Clades (Fig. 3A–F). The relationships among these clades, however, are not well resolved within individual gene trees.

3.4. MtDNA vs. nDNA trees

The RAxML concatenated analyses indicate several areas of discordance between the mtDNA and nDNA datasets (Fig. 4). Both trees show strong support for the Trilling Frog, Fat Frog, *crucifer*, and West Coast Clades. The concatenated mtDNA tree, however, does not support the relationships among these clades, with bootstrap values <60 for these deeper nodes. The nDNA tree strongly supports (bootstrap values = 100) a sister relationship between the West Coast Clade and all other *Pseudacris*, a sister relationship between the Fat Frog and *crucifer* Clades, and together these two clades form the sister group to the Trilling Frog Clade. Within each of the clades, the putative taxa are monophyletic with the exception of the *regilla*, *sierra*, and *hypochondriaca* group, the *streckeri* and *illinoensis* group, and the inland and coastal *feriarum* group (Fig. 4B).

We find one striking difference between the mtDNA and nDNA concatenated trees that suggests potential evidence of mitochondrial introgression. Within the Trilling Frog Clade, *P. clarkii* and *P. maculata* are not reciprocally monophyletic based on mtDNA (Fig. 4A). In the nDNA tree, each species forms a strongly supported (bootstrap value = 100) monophyletic group (Fig. 4B), although the placement of these species within the Trilling Frog Clade is unclear.

3.5. Concatenated and species trees

The concatenated nDNA tree (Fig. 4B) and the species trees estimated using BUCKy and *BEAST (Fig. 5) have similar overall topologies. The four major clades exhibit the same relationships in all analyses with the 26 nDNA loci combined: a sister relationship between the West Coast Clade and all other *Pseudacris*, a sister relationship between the Fat Frog and *crucifer* Clades, and a sister relationship between the Fat Frog/*crucifer* and the Trilling Frog Clades. Between the concatenated nDNA and the BUCKy trees, the only topological differences include the position of one of the *sierra* individuals (I28) and one of the inland *feriarum* individuals (I11), but these branches are weakly supported (bootstrap value < 75) and have low CFs (0.06–0.18).

We found two differences between the relationships estimated using the single-allele datasets in BUCKy and *BEAST (Fig. 5). Within the Trilling Frog Clade, the positions of *P. clarkii* and *P. maculata* are inconsistent, and these branches have low CFs (0.08–0.12). Within the West Coast Clade, the branching pattern of *regilla*, *sierra*, and *hypochondriaca* also differs. The BUCKy analyses produced the same results regardless of the *a priori* level of discordance value (α) used, and sub-sampling the dataset to remove missing data did not change the overall topology. The BUCKy tree demonstrates the lack of reciprocal monophyly within some of the putative species (*sierra*, *hypochondriaca*, *streckeri*, and *illinoensis*), and the low CFs indicate the high level of discordance among the sampled nDNA loci.

The multiple-allele "species" tree estimated in *BEAST (Fig. 6) is largely congruent with the single-allele *BEAST tree (Fig. 5B). The only difference is the placement of *P. maculata*, which is recovered as the sister taxon to *P. fouquettei* in the multiple-allele tree (Fig. 6). The multiple-allele *BEAST analysis with every individual set as a taxon (44 tips) failed to converge after one billion generations, so the maximum clade credibility tree is not shown.

4. Discussion

4.1. Utility of RRLs and parallel sequencing for non-model organisms

In this study we targeted anonymous nuclear loci previously developed for *Pseudacris* using RRL sequencing (Lemmon and Lemmon, 2012) and demonstrate the effectiveness of parallel tagged sequencing (PTS) on an Illumina MiSeq platform. Recent studies that have employed PTS have used the 454 platform because of the longer read lengths produced (Bybee et al., 2011b; Griffin et al., 2011; O'Neill et al., 2013; Puritz et al., 2012). By sonicating our pooled amplicons prior to sequencing paired 150 bp reads we were able to assemble full sequences for loci 437–710 bp in length. Even with our strict quality settings, we obtained a nearly complete data matrix. This success was probably largely due to the time spent quantifying and pooling amplicons at equal concentrations, which is less feasible for projects with more individuals and more loci. The post-PCR wet-laboratory costs for this study including sonication, library preparation, quantitation, and 1/3 Illumina MiSeq run was approximately $2083 and resulted in sequence data from which allelic phase could be determined. The comparable cost of Sanger sequencing assuming forward and reverse reactions for each amplicon, using our in-house rate at the FSU Sequencing Facility of $4 per reaction, would be approximately $10,560. This amount does not include the cost of cloning (both time and money) that would be necessary to obtain phased alleles.

Alternative approaches for collecting high-throughput sequence data for phylogenetic studies of non-model organisms continue to be developed. Several factors may need to be considered when choosing an approach for a given study including the previous genomic resources available, the size of the project (number of individuals and loci), and the experience of the researcher. The PTS approach may be appropriate for small to medium sized projects, e.g., 320 individuals and 5 loci (Puritz et al., 2012); 93 individuals and 95 loci (O'Neill et al., 2013); 44 individuals and 30 loci (this study). The approach is appealing because it uses tradional PCR approaches such that orthologous loci can be targeted using previously designed primers, but does not scale up well to larger projects because of PCR costs and the time necessary for pooling amplicons (especially without the use of a liquid-handling robot). A different approach, hybrid enrichment (also known as sequence capture), involves hybridizing probe sequences to genomic DNA to enrich for regions of interest prior to sequencing (Albert et al., 2007; Gnirke et al., 2009). This approach, in combination with probe sets designed for capture across broad taxonomic groups (Faircloth et al., 2012; Lemmon et al., 2012), is rapidly producing datasets with hundreds of loci for phylogenetic studies of non-model organisms (e.g., McCormack et al., 2013b). The initial start-up costs of probe design and purchase are not trivial, making these approaches more suitable for large-scale projects or for encouraging collaborations between laboratories. Additional approaches for collecting genomic data for phylogenetics have been reviewed in detail previously (Lemmon and Lemmon, 2013; McCormack et al., 2013a).

4.2. Mito-nuclear discordance

The discordance between the mtDNA and nDNA phylogenies estimated in this study can be attributed to at least two factors – different levels of information and differences likely due to introgression. The mtDNA loci we sequenced contain more variable and informative sites than any of the individual anonymous nuclear loci. It is the high mutation rate and small effective population size of mtDNA that have made it the data of choice for many phylogenetic and phylogeographic studies (Avise et al., 1987; Wilson
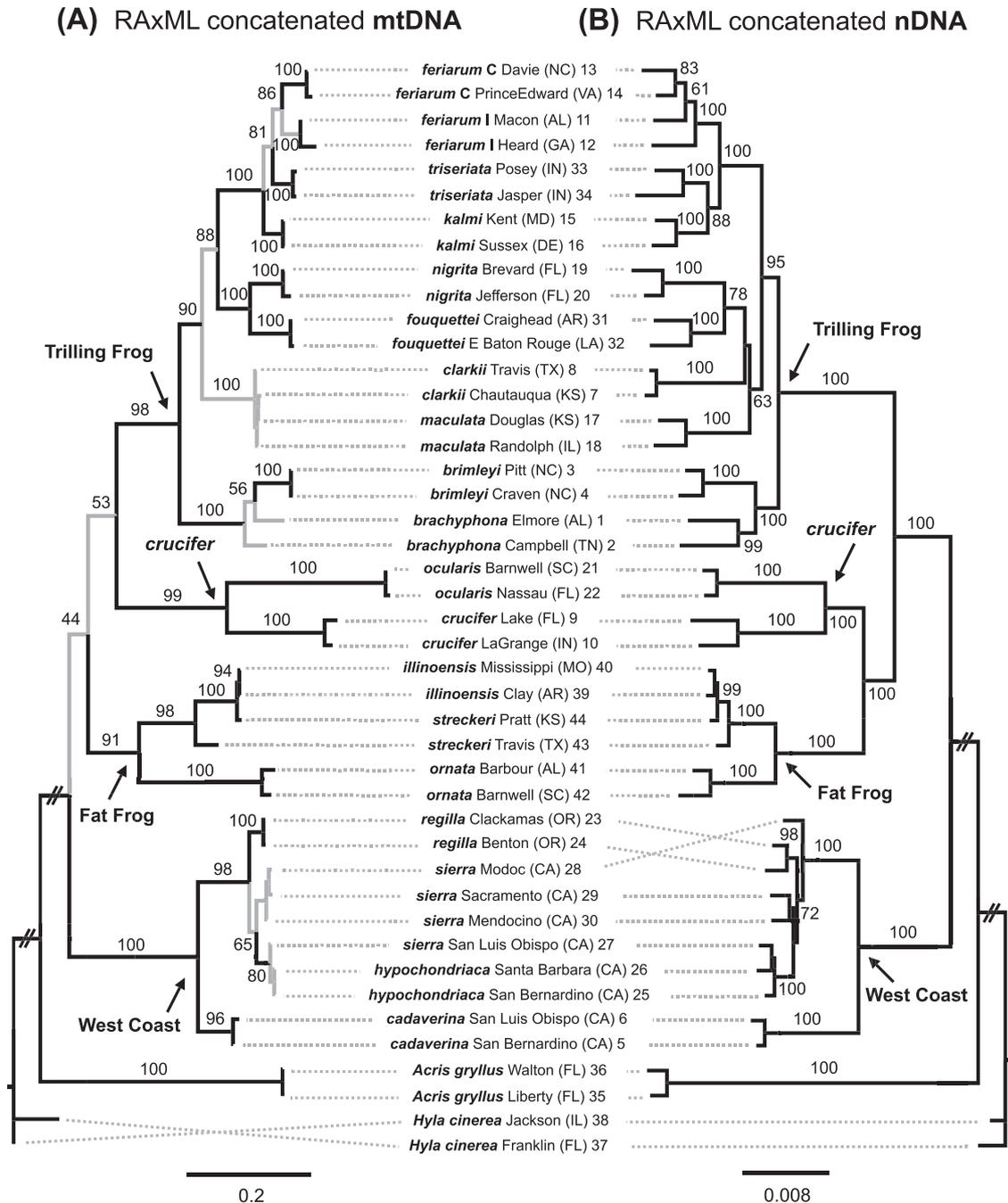
Fig. 4. Concatenated RAxML trees estimated with (A) three mtDNA loci (1886 total bp with 656 ingroup informative sites) and (B) 26 nDNA loci (12,679 total bp with 1285 ingroup informative sites). Branch labels are bootstrap support values based on 1000 replicates and scale bars correspond to the mean number of nucleotide substitutions per site. Gray branches indicate areas of disagreement between mtDNA and nDNA.

et al., 1985). The nDNA loci we sequenced, however, enabled us to resolve the deeper nodes in the tree, further highlighting the need for multilocus data, and perhaps loci with different levels of information in order to resolve all parts of the tree. In our case, the combined nDNA data resolved the shallow nodes as well, so we chose not to combine the mtDNA and nDNA into a single analysis because of the obvious discrepancies in the mtDNA tree that we do not believe reflect the overall species history.

Mito-nuclear discordance has been noted in a number of systems, and introgression and incomplete lineage sorting are two common explanations (Toews and Brelsford, 2012). Lemmon et al. (2007a) sequenced 12S/16S mtDNA with a much wider geographic sampling of *P. clarkii* and *P. maculata* and found that they

share mtDNA throughout their ranges. Taken together with the patterns observed in our nDNA data, in which *P. clarkii* and *P. maculata* form distinct monophyletic groups, these findings seem most consistent with introgression and a potential selective sweep of mtDNA. Additional geographic sampling with nuclear loci will be necessary to further elucidate these patterns. Because mtDNA effectively represents only a single history, it may be sensible to remove it from species tree analyses when these conflicts occur. Fortunately, we can be more selective about the data we include now that it is feasible to collect many independent nuclear loci with sufficient levels of variation, although development of methods for sub-sampling data in an unbiased manner is still in the early stages.
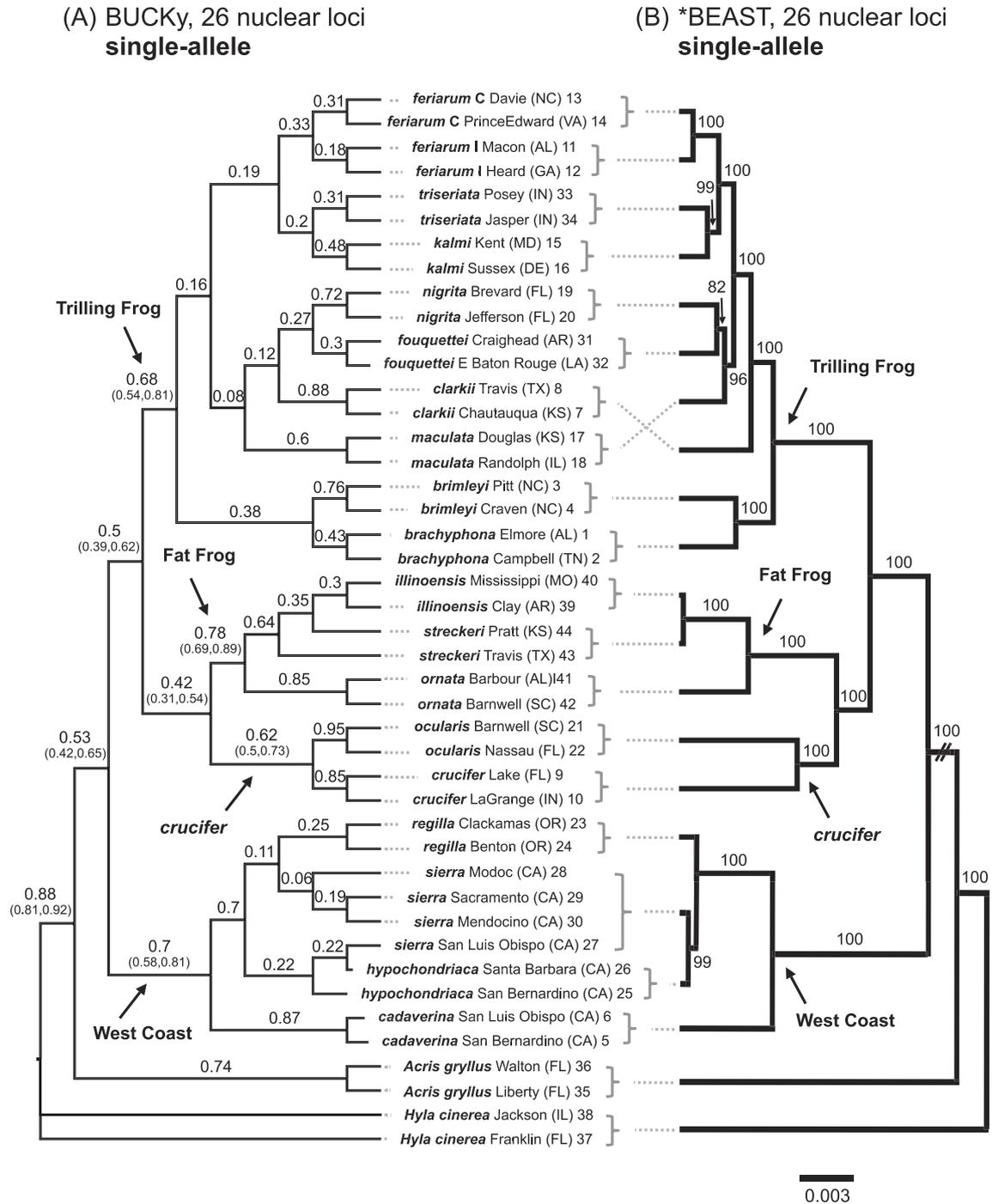
Fig. 5. Species trees estimated with (A) BUCKy and (B) *BEAST using the single-allele dataset (26 nDNA loci). Branch labels represent (A) sample-wide Concordance Factors with 95% credibility intervals shown for the major clades and (B) Bayesian posterior probabilities. In (B), tips were set as the putative species/lineage of each individual, shown in bold italics in the taxon labels.

4.3. Comparing species tree methods

Although there was a high level of discordance among our nDNA gene trees, we estimated largely similar topologies within *Pseudacris* using concatenated RAxML, BUCKy, and *BEAST analyses. The relationships among the major clades were consistent across methods, and the discrepancies occurred in the more shallow nodes. One major concern with concatenation is that high discordance among genes can produce an incorrect tree that is strongly supported (Belfiore et al., 2008; Degnan and Rosenberg, 2006; Edwards et al., 2007; Kubatko and Degnan, 2007; Weisrock et al., 2012). This issue highlights the need to compare across

methods with different underlying assumptions in order to gain more confidence in the resulting tree. In particular, methods that incorporate more biological realism by addressing the discordance among gene trees are perhaps the most appropriate for species tree estimation.

The primary concordance tree we estimated using BUCKy demonstrates the areas in the tree that have low concordance across the sampled loci. Although this is not a "species" tree *per se*, we find largely similar relationships when compared to the *BEAST tree. The only differences are the placement of *P. maculata* and *P. clarkii*, and the branching pattern of *regilla*, *sierra*, and *hypochondriaca*. Given that *BEAST does not account for hybridization or gene

Fig. 6. Species tree estimated with *BEAST using the multiple-allele dataset (26 nDNA loci). Branch labels represent Bayesian posterior populations. Tips were set as the putative species/lineage of each individual.

flow, and the low concordance factors in the BUCKy tree indicate disagreement among loci at these nodes, the results are not surprising. One challenge of applying high-throughput sequencing methods to phylogenetics is the computational burden of analyzing massive datasets. When we analyzed our multiple-allele dataset with 26 nDNA loci and 44 tips in *BEAST, the runs did not converge after one billion generations with relatively simple parameter settings (e.g., strict clock, constant population size). Similar problems with convergence have occurred in other studies that used coalescent-based methods with multiple-allele datasets, e.g., 18 loci with 26 tips in BEST (Lee et al., 2012) and 50–94 loci with 9 tips in *BEAST (O'Neill et al., 2013). When O'Neill et al. (2013) used subsets of their most informative loci (up to 50 loci) the analyses successfully converged, further highlighting the need to focus on the data underlying phylogenetic signal and filter out the noise (Hillis and Huelsenbeck, 1992; Jeffroy et al., 2006; Philippe et al., 2005; Rodríguez-Ezpelata et al., 2007; Townsend, 2007).

4.4. Evolutionary relationships within Pseudacris

The multilocus nuclear data we obtained through parallel tagged amplicon sequencing allowed us to elucidate the relationships among the major clades of *Pseudacris* chorus frogs, which were previously ambiguous. Our analyses of 26 anonymous nuclear loci corroborate the monophyly of four clades (Trilling Frog, Fat Frog, *crucifer*, and West Coast) previously described based on mtDNA (Lemmon et al., 2007a; Moriarty and Cannatella, 2004). Our nDNA dataset also clarified (1) the sister relationship between the West Coast Clade and all other *Pseudacris*, (2) a sister relationship between the Fat Frog and *crucifer* Clades, and (3) a sister relationship between the Fat Frog/*crucifer* and Trilling Frog Clades. This generally well-supported phylogeny provides a framework for future studies investigating behavioral evolution and the timing of divergence in this genus.

Within the Trilling Frog Clade, we find consistent support for the sister relationship between *P. brachyphona* and *P. brimleyi* and their sister relationship to all other members of the group. Our species tree analyses also indicate support for two reciprocally monophyletic clades of *P. feriarum* (an inland and coastal clade), but further geographic sampling with nDNA loci is needed. Interestingly, we find that the nDNA data supports a sister relationship between *P. triseriata* and *P. kalmi*, in contrast to the mtDNA results from Lemmon et al. (2007a). This finding lends support to the hypothesis proposed by Smith (1957) that *P. kalmi* represent relict populations of *P. triseriata* left by an expansion and contraction event following the Wisconsin glaciation.

Our analyses suggest a few areas of the *Pseudacris* phylogeny that require additional investigation. We demonstrate that *P. maculata* and *P. clarkii* differentiate based on nDNA data, which is consistent with morphological and acoustic data, despite sharing mtDNA throughout their ranges (Lemmon et al., 2007a). Their positions within the Trilling Frog Clade, however, are remarkably inconsistent among analyses, even between the single-allele and multiple-allele *BEAST analyses. One possible explanation is that the methods we used do not explicitly account for hybridization. It may be informative to further explore this issue with phylogenetic network analyses that model reticulation events such as those implemented in STEM-hy (Kubatko, 2009) or PhyloNet (Than et al., 2008; Yu et al., 2011). It is also unclear what effect missing data may have on species tree analyses, although our BUCKy analyses with and without missing data produced consistent topologies. Both our mtDNA and nDNA datasets suggest that *P. streckeri* is paraphyletic with respect to *P. illinoensis*, and that *P. regilla, P. sierra,* and *P. hypochondriaca* do not form well-supported monophyletic groups. In both cases, putative species have possibly undergone recent divergences, resulting in incomplete lineage sorting. Our results, though limited in geographic scope, do not support the species designations suggested by Recuero et al. (2006) for mtDNA lineages of the *P. regilla* complex. Further

geographic sampling that incorporates behavioral data with multilocus genetic data may be necessary to clarify the taxonomy and range boundaries of these groups.

5. Conclusions

We demonstrate the utility of RRL sequencing for developing anonymous nuclear loci for phylogenetic studies in non-model organisms. We successfully amplified and sequenced 27 of these nuclear loci along with three mitochondrial loci for 44 individuals on one-third of an Illumina MiSeq run. Our approach resulted in a data matrix that is 96.0% complete at less than 20% of the cost of traditional Sanger sequencing. We find evidence of mito-nuclear discordance most likely due to introgression, and species complexes that have possibly undergone recent divergences. These taxa require further investigation with wider geographic sampling and multilocus datasets in order to delimit taxonomic and geographic boundaries. Our species tree analyses with 26 nuclear loci fully resolve the relationships among the major clades of *Pseudacris* chorus frogs, providing a new framework for future studies of behavioral evolution and speciation in this group.

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.2014.02.007.

References

Albert, T.J., Molla, M.N., Muzny, D.M., Nazareth, L., Wheeler, D., Song, X., Richmond, T.A., Middle, C.M., Rodesch, M.J., Packard, C.J., Weinstock, G.M., Gibbs, R.A., 2007. Direct selection of human genomic loci by microarray hybridization. Nat. Methods 4, 903–905.

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., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A., Saunders, N.C., 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu. Rev. Ecol. Syst. 18, 489–522.

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

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

Bybee, S.M., Bracken-Grissom, H., Haynes, B.D., Hermansen, R.A., Byers, R.L., Clement, M.J., Udall, J.A., Wilcox, E.R., Crandall, K.A., 2011a. Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics. Genome Biol. Evol. 3, 1312–1323.

Bybee, S.M., Bracken-Grissom, H.D., Hermansen, R.A., Clement, M.J., Crandall, K.A., Felder, D.L., 2011b. Directed next generation sequencing for phylogenetics: an example using Decapoda (Crustacea). Zool. Anz. 250, 497–506.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552.

Chung, Y., Ané, C., 2011. Comparing two Bayesian methods for gene tree/species tree reconstructions: simulations with incomplete lineage sorting and horizontal gene transfer. Syst. Biol. 60, 261–275.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., Blaxter, M.L., 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. 12, 499–510.

Degnan, J.H., Rosenberg, N.A., 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2, e68.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

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

Edwards, S.V., Liu, L., Pearl, D.K., 2007. High-resolution species trees without concatenation. Proc. Natl. Acad. Sci. U.S.A. 104, 5936–5941.

Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. 61, 717–726.

Fouquette, M.J., 1975. Speciation in chorus frogs I. Reproductive character displacement in the *Pseudacris nigrita* complex. Syst. Zool. 24, 16–22.

Gartside, D.F., 1980. Analysis of a hybrid zone between chorus frogs of the *Pseudacris nigrita* complex in the southern United States. Copeia 1, 56–66.

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E.M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D.B., Lander, E.S., Nusbaum, C., 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat. Biotechnol. 27, 182–189.

Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. 27, 221–224.

Griffin, P.C., Robin, C., Hoffmann, A.A., 2011. A next-generation sequencing method for overcoming the multiple gene copy problem in polyploidy phylogenetics, applied to *Poa* grasses. BMC Biol. 9, 19.

Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27, 570–580.

Hillis, D.M., Huelsenbeck, J.P., 1992. Signal, noise, and reliability in molecular phylogenetic analyses. J. Hered. 83, 189–195.

IUCN, 2012. IUCN Red List of Threatened Species. Version 2012.2. <www.iucnredlist.org> (downloaded 13.02.13).

Jeffroy, O., Brinkmann, H., Delsuc, F., Philippe, H., 2006. Phylogenomics: the beginning of incongruence? Trends Genet. 22, 225–231.

Knowles, L.L., 2009. Statistical phylogeography. Annu. Rev. Ecol. Evol. Syst. 40, 593–612.

Kubatko, L.S., 2009. Identifying hybridization events in the presence of coalescence via model selection. Syst. Biol. 58, 478–488.

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

Kubatko, L.S., Carstens, B.C., Knowles, L.L., 2009. STEM: species tree estimation using maximum likelihood for gene trees under coalescence. Bioinformatics 25, 971–973.

Larget, B.R., Kotha, S.K., Dewey, C.N., Ané, C., 2010. BUCKy: gene tree/species tree reconciliation with Bayesian concordance analysis. Bioinformatics 26, 2910–2911.

Leaché, A.D., Rannala, B., 2011. The accuracy of species tree estimation under simulation: a comparison of methods. Syst. Biol. 60, 126–137.

Lee, J.Y., Joseph, L., Edwards, S.V., 2012. A species tree for the Australo-Papuan fairy-wrens and allies (Aves: Maluridae). Syst. Biol. 61, 253–271.

Lemmon, E.M., 2009. Diversification of conspecific signals in sympatry: geographic overlap drives multidimensional reproductive character displacement in frogs. Evolution 63, 1155–1170.

Lemmon, E.M., Lemmon, A.R., 2010. Reinforcement in chorus frogs: lifetime fitness estimates including intrinsic natural selection and sexual selection against hybrids. Evolution 64, 1748–1761.

Lemmon, A.R., Lemmon, E.M., 2012. High-throughput identification of informative nuclear loci for shallow-scale phylogenetics and phylogeography. Syst. Biol. 61, 745–761.

Lemmon, E.M., Lemmon, A.R., 2013. High-throughput genomic data in systematics and phylogenetics. Annu. Rev. Ecol. Evol. Syst. 44, 99–121.

Lemmon, E.M., Lemmon, A.R., Collins, J.T., Lee-Yaw, J.A., Cannatella, D.C., 2007a. Phylogeny-based delimitation of species boundaries and contact zones in the trilling chorus frogs (*Pseudacris*). Mol. Phylogenet. Evol. 44, 1068–1082.

Lemmon, E.M., Lemmon, A.R., Cannatella, D.C., 2007b. Geological and climatic forces driving speciation in the continentally distributed trilling chorus frogs (*Pseudacris*). Evolution 61, 2086–2103.

Lemmon, A.R., Emme, S.A., Lemmon, E.M., 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst. Biol. 61, 727–744.

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

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

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

Mamanova, L., Coffey, A.J., Scott, C.E., Kozarewa, I., Turner, E.H., Kumar, A., Howard, E., Shendure, J., Turner, D.J., 2010. Target-enrichment strategies for next-generation sequencing. Nat. Methods 7, 111–118.

McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C., Brumfield, R.T., 2013a. Applications of next-generation sequencing to phylogeography and phylogenetics. Mol. Phylogenet. Evol. 66, 526–538.

McCormack, J.E., Harvey, M.G., Faircloth, B.C., Crawford, N.G., Glenn, T.C., Brumfield, R.T., 2013b. A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. PLoS ONE 8, e54848.

Meyer, M., Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. 2010. http://dx.doi.org/10.1101/pdb.prot5448.

Moriarty, E.C., Cannatella, D.C., 2004. Phylogenetic relationships of the North American chorus frogs (*Pseudacris*: Hylidae). Mol. Phylogenet. Evol. 30, 409–420.

Neill, W.T., 1949. The status of Baird's chorus-frog. Copeia 3, 227–228.

Nylander, J.A.A., 2004. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

O'Neill, E.M., Schwartz, R., Bullock, C.T., Williams, J.S., Shaffer, H.B., Aguilar-Miguel, X., Parra-Olea, G., Weisrock, D.W., 2013. Parallel tagged amplicon sequencing reveals major lineages and phylogenetic structure in the North American tiger salamander (*Ambystoma tigrinum*) species complex. Mol. Ecol. 22, 111–129.

Pamilo, P., Nei, M., 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5, 568–583.

Philippe, H., Delsuc, F., Brinkmann, H., Lartillot, N., 2005. Phylogenomics. Annu. Rev. Ecol. Evol. Syst. 36, 541–562.

Platz, J.E., 1989. Speciation within the chorus frog *Pseudacris triseriata*: morphometric and mating call analyses of the boreal and Western subspecies. Copeia 3, 704–712.

Platz, J.E., Forester, D.C., 1988. Geographic variation in mating call among the four subspecies of the chorus frog: *Pseudacris triseriata* (Wied). Copeia 4, 1062–1066.

Puritz, J.B., Addison, J.A., Toonen, R.J., 2012. Next-generation phylogeography: a targeted approach for multilocus sequencing of non-model organisms. PLoS ONE 7, e34241.

Rambaut, A., Drummond, A.J., 2007. Tracer v1.4, <http://beast.bio.ed.ac.uk/Tracer>.

Recuero, E., Martínez-Solano, Í., Parra-Olea, G., García-París, M., 2006. Phylogeography of *Pseudacris regilla* (Anura: Hylidae) in western North America, with a proposal for a new taxonomic rearrangement. Mol. Phylogenet. Evol. 39, 293–304.

Rodríguez-Ezpelata, N., Brinkmann, H., Roure, B., Lartillot, N., Lang, B.F., Philippe, H., 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. Syst. Biol. 56, 389–399.

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

Slowinski, J.B., Page, R.D.M., 1999. How should species phylogenies be inferred from sequence data? Syst. Biol. 48, 814–825.

Smith, P.W., 1957. An analysis of post-Wisconsin biogeography of the Prairie Peninsula region based on distributional phenomena among terrestrial vertebrate populations. Ecology 38, 205–218.

Smith, P.W., Smith, D.M., 1952. The relationship of the chorus frogs, *Pseudacris nigrita feriarum* and *Pseudacris n. triseriata*. Am. Midl. Nat. 48, 165–180.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.

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

Talavera, G., Castresana, J., 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56, 564–577.

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

Toews, D.P.L., Brelsford, A., 2012. The biogeography of mitochondrial and nuclear discordance in animals. Mol. Ecol. 21, 3907–3930.

Townsend, J.P., 2007. Phylogenetic informativeness. Syst. Biol. 56, 222–231.

Weisrock, D.W., Smith, S.D., Chan, L.M., Biebouw, K., Kappeler, P.M., Yoder, A.D., 2012. Concatenation and concordance in the reconstruction of mouse lemur phylogeny: an empirical demonstration of the effect of allele sampling in phylogenetics. Mol. Biol. Evol. 29, 1615–1630.

Wilson, A.C., Cann, R.L., Carr, S.M., George, M., Gyllensten, U.B., Helm-Bychowski, K.M., Higuchi, R.G., Palumbi, S.R., Prager, E.M., Sage, R.D., Stoneking, M., 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. Biol. J. Linn. Soc. Lond. 26, 375–400.

Yu, Y., Than, C., Degnan, J.H., Nakhleh, L., 2011. Coalescent histories on phylogenetic networks and detection of hybridization despite incomplete lineage sorting. Syst. Biol. 60, 138–149.