A Versatile and Highly Efficient Toolkit Including 102 Nuclear Markers for Vertebrate Phylogenomics, Tested by Resolving the Higher Level Relationships of the Caudata

Xing Xing Shen,[1] Dan Liang,[1] Yan Jie Feng,[1] Meng Yun Chen,[1] and Peng Zhang*[,1]

[1]Key Laboratory of Gene Engineering of the Ministry of Education, State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-Sen University, Guangzhou, China

***Corresponding author:** E-mail: alarzhang@gmail.com.

Associate editor: Xun Gu

Abstract

Resolving difficult nodes for any part of the vertebrate tree of life often requires analyzing a large number of loci. Developing molecular markers that are workable for the groups of interest is often a bottleneck in phylogenetic research. Here, on the basis of a nested polymerase chain reaction (PCR) strategy, we present a universal toolkit including 102 nuclear protein-coding locus (NPCL) markers for vertebrate phylogenomics. The 102 NPCL markers have a broad range of evolutionary rates, which makes them useful for a wide range of time depths. The new NPCL toolkit has three important advantages compared with all previously developed NPCL sets: 1) the kit is universally applicable across vertebrates, with a PCR success rate of 94.6% in 16 widely divergent tested vertebrate species; 2) more than 90% of PCR reactions produce strong and single bands of the expected sizes that can be directly sequenced; and 3) all cleanup PCR reactions can be sequenced with only two specific universal primers. To test its actual phylogenetic utility, 30 NPCLs from this toolkit were used to address the higher level relationships of living salamanders. Of the 639 target PCR reactions performed on 19 salamanders and several outgroup species, 632 (98.9%) were successful, and 602 (94.1%) were directly sequenced. Concatenation and species-tree analyses on this 30-locus data set produced a fully resolved phylogeny and showed that Cryptobranchoidea (Cryptobranchidae + Hynobiidae) branches first within the salamander tree, followed by Sirenidae. Our experimental tests and our demonstration for a particular case show that our NPCL toolkit is a highly reliable, fast, and cost-effective approach for vertebrate phylogenomic studies and thus has the potential to accelerate the completion of many parts of the vertebrate tree of life.

Key words: **nuclear marker, phylogenomic, vertebrate, salamander, phylogeny.**

Introduction

Building phylogenomic supermatrices with multiple nuclear loci has become the standard method of resolving species relationships in difficult biological scenarios (Delsuc et al. 2005). One efficient method of constructing multilocus data sets is expressed sequence tag (EST) (Philippe and Telford 2006; Dunn et al. 2008; Philippe et al. 2009) or transcriptome (Künstner et al. 2010) sequencing, in which high-quality RNA is extracted from each organism of interest and a huge number of ESTs or transcripts are then sequenced by Sanger or next-generation sequencing (NGS). However, this approach often generates patchy data sets with a high proportion of missing data, which may compromise phylogenetic inference (Lemmon et al. 2009; Roure et al. 2013). More importantly, this approach is not workable for many older collections because these specimens can only provide DNA samples. A second efficient way to construct multilocus data sets is the sequence capture method in which target genomic regions are selectively captured by hybridization with probes before NGS (Crawford et al. 2012; Faircloth et al. 2012; Lemmon et al. 2012; McCormack et al. 2012). The most attractive feature of this method is that it can generate hundreds to thousands of loci for many samples in a short time. However, experimentally, the efficiency of sequence capture is considerably influenced by the divergence between probes and target sequences (Lemmon et al. 2012; McCormack et al. 2013). More importantly, turning the huge data set derived from sequence capture into sequences that researchers can analyze requires sophisticated bioinformatic processing, which is currently quite challenging to most phylogenetic researchers (McCormack et al. 2013). Therefore, although the sequence capture method is efficient and promising, its immaturity currently restricts its wide application in the community.

Currently, for vertebrate phylogenetics, the most widely used approach for building multilocus data sets is still conventional targeted polymerase chain reaction (PCR) and the sequencing of selected orthologous genes. However, the PCR-based method is laborious: 1) most practitioners spend much time developing and screening molecular markers that are workable for their studied taxa and suitable to their evolutionary timescale of interest (Murphy et al. 2001; Li et al. 2007; Townsend et al. 2008; Wright et al. 2008; Shen et al. 2011); 2) it requires PCR of each organism at

Article

Fast Track

each locus, not to mention the extra effort involved in PCR optimization, gel-purification, and cloning. On the other hand, the PCR-based method also has its advantages: 1) it is highly targeted and can produce nearly complete data matrices, and the data analysis process is straightforward and familiar to most empirical researchers; 2) it requires no prior genomic knowledge of the targeted organisms; and 3) it works with tiny amounts of DNA and thus appears to be an ideal solution when DNA samples are limited.

For most interspecific phylogenetic projects, nuclear protein-coding loci (NPCLs) that are developed on exons are likely the markers of choice for the PCR-based strategy, because they provide an appropriate level of variation, easy alignment across a large phylogenetic span, and relatively straightforward detection of paralogs (Thomson et al. 2010). In this study, our aim was to develop a suite of universal NPCL markers and an efficient experimental protocol for vertebrate phylogenomics. Aimed at eliminating the drawbacks of the conventional PCR-based method, we designed our NPCL toolkit and protocol to 1) include approximately 100 NPCL markers (we think the economic transition from PCR to sequence capture is at approximately 100 loci; if more than 100 loci are to be used, the PCR method is not cost-efficient); 2) work for all major jawed vertebrate clades and provide good resolution at different evolutionary timescales; 3) produce single and strong amplicon bands without any PCR optimization in most cases; and 4) yield PCR products that can be directly cleaned and sequenced without gel purification or cloning in most cases.

Because our NPCL toolkit is designed for universal phylogenetic applications in vertebrates, it should be tested in a real case with some difficult samples. Salamanders are well known to have much larger genomes than most vertebrates (often 10 times the human genome, http://www.genomesize.com/). The PCR-based method normally performs poorly for salamanders (personal experience and communication with colleagues). For example, Shen et al. (2011) amplified 22 NPCL markers in 16 tested vertebrates. In all 15 nonsalamander species, approximately 90% of the markers could be successfully amplified; however, for the tested salamander species *Batrachuperus yenyuanensis*, only 8 of 22 NPCL markers (36%) could be amplified. Here, we apply our NPCL toolkit and protocol to address the higher level relationships of living salamanders as a test of the toolkit's utility. Our results demonstrate that the new universal NPCL toolkit and protocol are fast and effective in constructing multilocus data matrices for vertebrate phylogenomics.

Results

Experimental Performance and Characteristics of the New NPCL Toolkit

The newly developed NPCL toolkit contains 102 NPCL markers, ranging from 510 to 1,650 bp, with an average length of 1,050 bp; each NPCL marker comprises two pairs of primers for the nested PCR strategy (supplementary table S1, Supplementary Material online). These 102 NPCL markers are broadly distributed on 21 chromosomes of the human genome (fig. 1). We classified their PCR performance into three levels: 1) producing a single target band of the expected size, 2) producing a target band but also significant nonspecific bands, and 3) not producing a target band. The first two conditions are considered successful. The PCR performances of the 102 NPCL markers across 16 diverse vertebrate species and three representative electropherograms are shown in figure 2. Of the 102 NPCL markers, 57 have a 100% success rate in the 16 tested vertebrate species, 87 have a success rate of more than 90%, and the remaining 15 range from 56% to 88% (fig. 2). Of the 1,632 PCR reactions (102 loci × 16 taxa), 1,544 (94.6%) were successful, with 1,485 (91%) producing strong, single target bands that can be used for direct sequencing. In the demonstration case in which 30 NPCL markers were used to investigate the higher level relationships of living salamanders, 632 (98.9%) of the 639 target fragments were successful. Of the 632 successful reactions, 602 (95%) were directly sequenced with the general sequencing primers "Seq_F" and "Seq_R." The PCR success rates for each of the 102 NPCL markers across the 16 tested vertebrate species are shown in figure 3.

The evolutionary rate, as evidenced by the degree of variability, is an important parameter of an NPCL marker because it determines applicability for different phylogenetic questions. Although our NPCL toolkit has a high PCR success rate in highly diverged taxa, that success does not mean that the NPCL markers in the toolkit are very conserved. As figure 3 illustrates, our toolkit includes NPCLs with a broad range of evolutionary rates, approximately 4-fold. Among the 102 NPCL markers, 60 evolve faster than RAG1, an NPCL that has been widely used for phylogenetic inference in various vertebrate groups. Because previous analyses based on RAG1 data resulted in highly resolved and robustly supported phylogenetic relationships at multiple hierarchical levels (San Mauro et al. 2005; Wiens et al. 2005; Hugall et al. 2007; Roelants et al. 2007), this indicates that our NPCL toolkit has the potential to resolve questions of both deep and shallow phylogeny.

It is well known that the fish-specific genome duplication occurred in the teleosts (Meyer and Van de Peer 2005). Although most duplicated genes were secondarily lost, some were retained or evolved new functions. For an NPCL marker, if there are two similar copies in teleost genomes, it is difficult to check the orthologous status of the obtained fragments. To this end, we took the zebrafish sequence of each NPCL to Basic Local Alignment Search Tool (Blast) against all available teleost genomes in the ENSEMBL database. If an NPCL receives more than two Blast hits and the top Blast score is not more than twice the second Blast score, that NPCL might have an extra copy in teleost genomes. Using this method, of the 102 NPCLs, it was found that only six (CXCR4, GLCE, KCNF1, LINGO1, NTN1, and PCDH10) may have extra copies in teleost genomes (fig. 3; supplementary table S1, Supplementary Material online). This result indicates that our NPCL toolkit is also suitable for phylogenetic inference in teleosts.

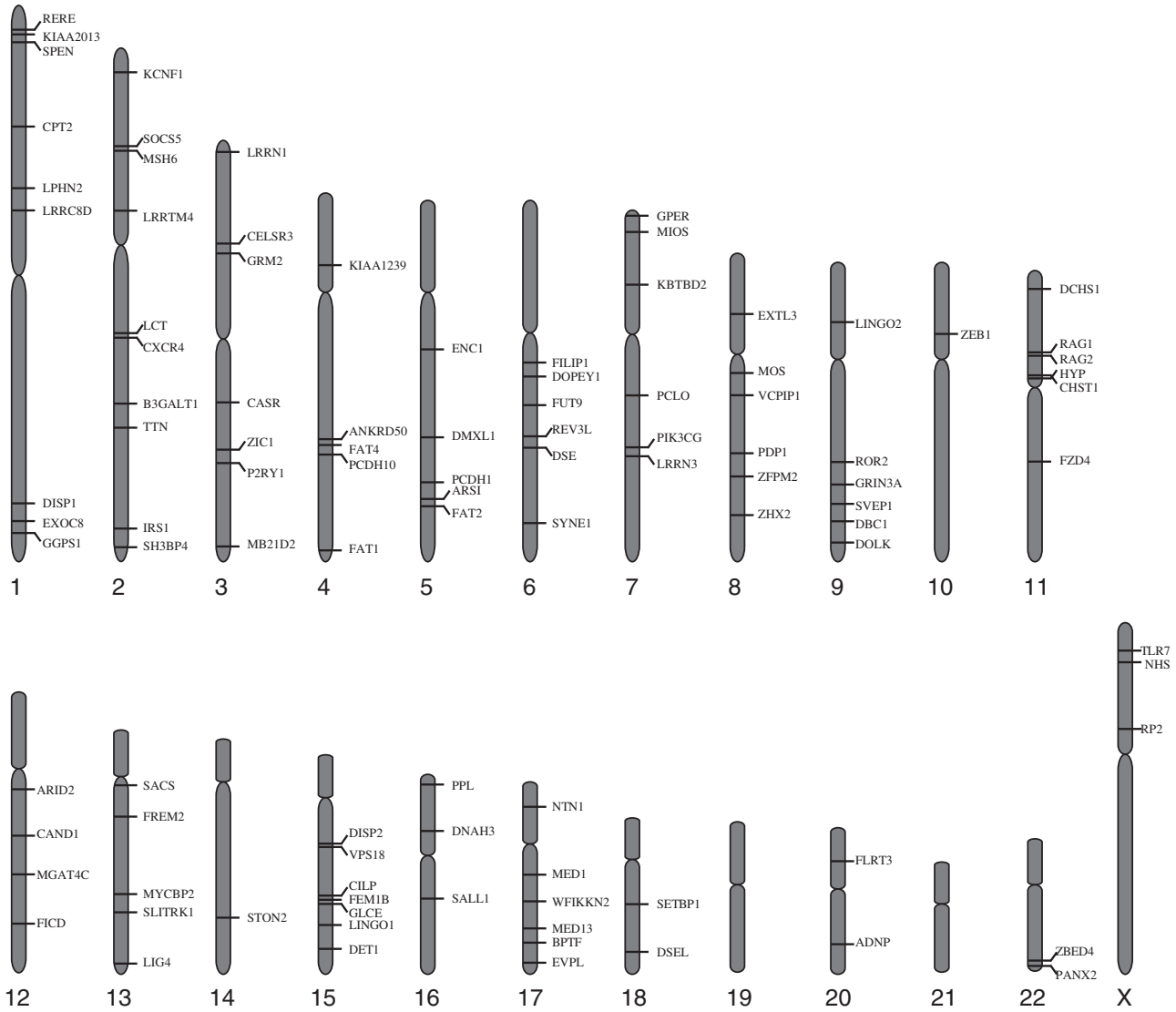

Fig. 1. Chromosome mapping of the 102 NPCL markers in the *Homo sapiens* genome.

Phylogenetic Performance in a Real Case

Our demonstration case included 19 salamander species that span salamander evolutionary diversity (supplementary table S2, Supplementary Material online). The nine outgroup species (two frogs, two caecilians, one turtle, one bird, two mammals, and one coelacanth) provided a largely balanced representation of relatives of salamanders. The 30 newly amplified NPCLs exhibited levels of variation comparable with that of the traditional RAG1 gene, with variable sites varying between 30% and 51% of all sequenced sites (table 1). The data set combining these 30 NPCLs comprises 27,834 bp and exhibits little substitution saturation (supplementary fig. S1, Supplementary Material online). The phylogenetic analyses of the concatenated data set using three tree-building methods (maximum likelihood [ML], Bayesian, and CAT-mixture model) produced an identical, fully resolved tree for 28 taxa (fig. 4). In all 25 nodes of the tree, the statistical support was highly robust (BP_{ML} 99–100%; $PP_{BAY} = 1.0$; $PP_{CAT} = 1.0$). The species tree estimated from 30 individual NPCLs without data concatenation using the pseudo-ML approach is identical to those estimated from the concatenated analyses. All nodes received bootstrap support values varying between 74 and 100 (fig. 4). We also conducted phylogenetic analyses at the amino acid level (9,278 deduced amino acid residues) using three tree-building methods (ML, Bayesian, and CAT-mixture model). The protein tree topology is identical to the DNA result with just slightly lower branch support for some nodes (supplementary fig. S2, Supplementary Material online). Therefore, we did not further analyze the protein data set.

The monophyly of extant amphibians with respect to amniotes and the close relationship between frogs and salamanders (the Batrachia hypothesis) are repeatedly recovered in most recent molecular studies (San Mauro et al. 2005; Zhang et al. 2005; Frost et al. 2006; Hugall et al. 2007; Roelants et al. 2007; Zhang and Wake 2009; San Mauro 2010; Pyron and Wiens 2011). However, a recent molecular study based on 26 nuclear genes (Fong et al. 2012) supports a caecilian–salamander sister relationship, with the possible paraphyly of extant amphibians. Our phylogenetic analyses based on

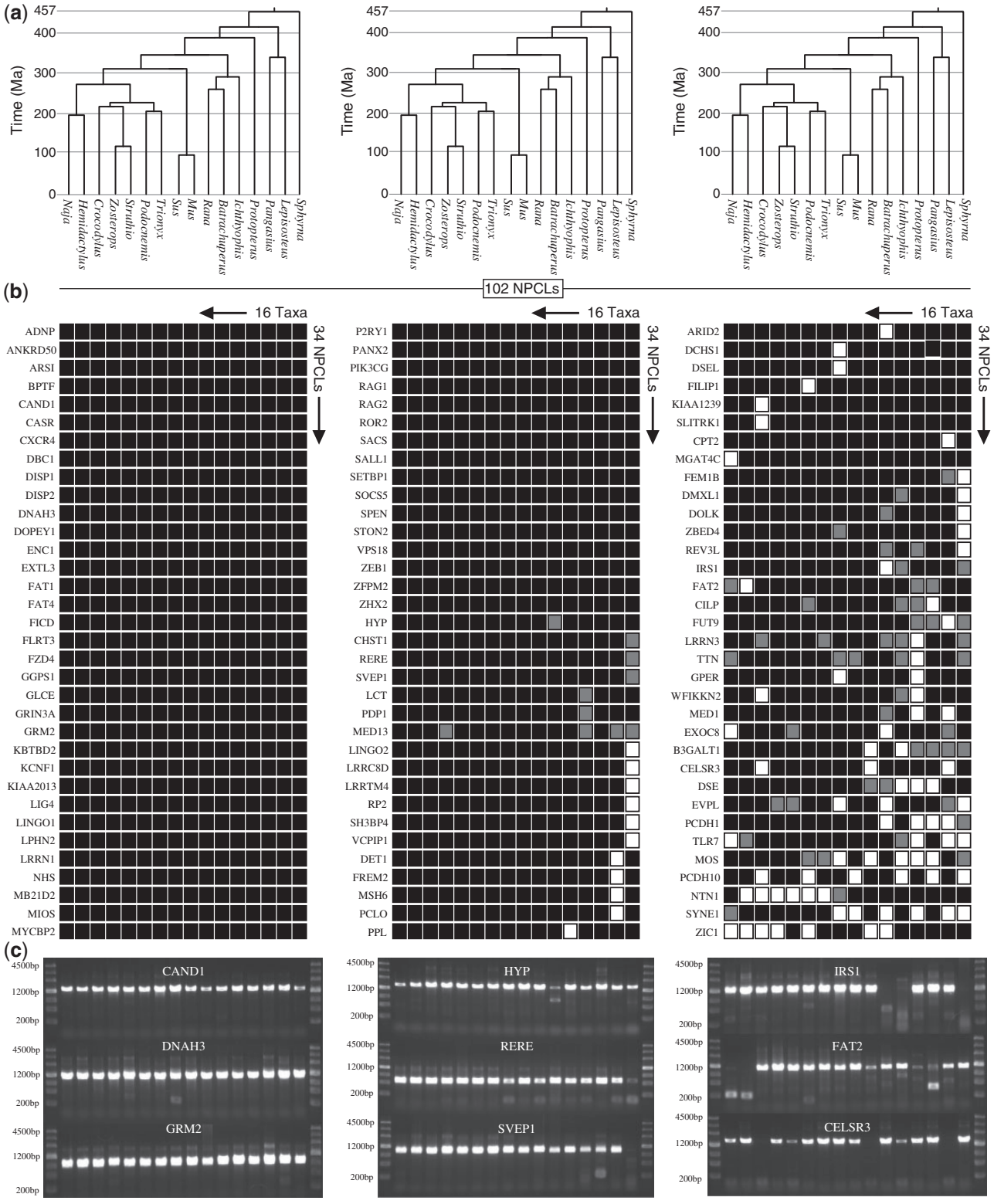

Fig. 2. PCR performance of the 102 NPCL markers in 16 divergent vertebrate species. Each square and electrophoretic lane is aligned with the tested species. (*a*) The draft divergence timescale for the 16 tested vertebrate species is based on Inoue et al. (2010) and the book *The Timetree of Life*. (*b*) The PCR performance of each NPCL marker is ranked by three different-colored squares: black, producing single target band; gray, having a target band but with significant nonspecific bands; white, no target band. The102 NPCL markers are sorted according to their PCR success rates. (*c*) Three typical agarose electrophoresis results for 9 NPCL markers.

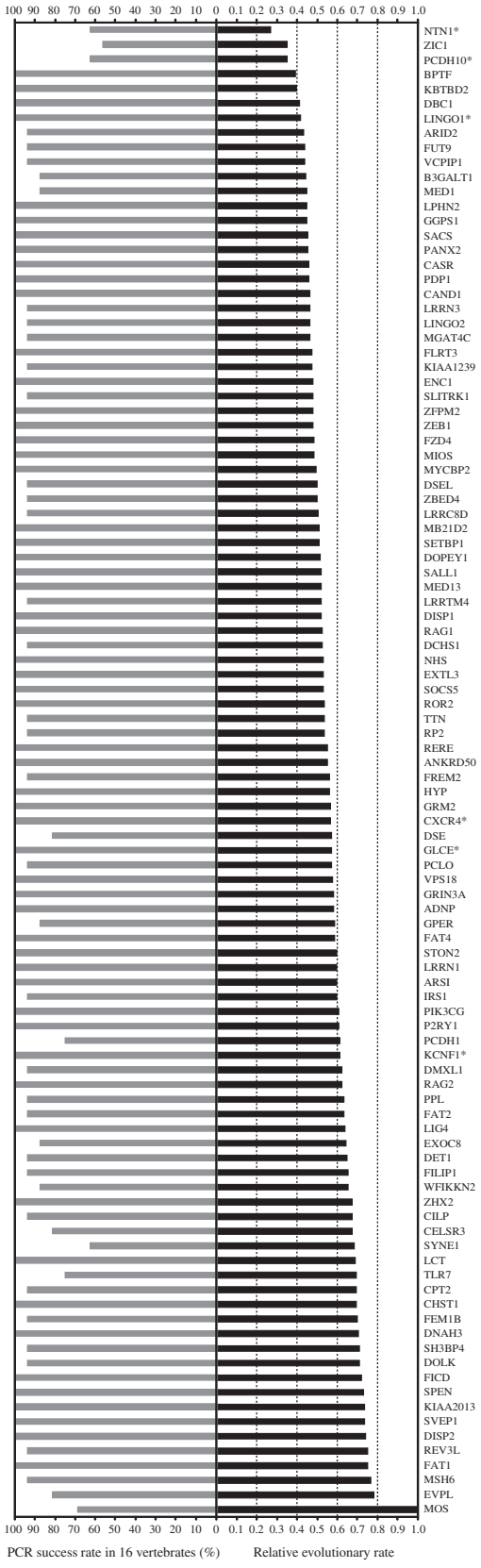

Fig. 3. Relative evolutionary rates of 102 NPCL in vertebrates. The 102 NPCLs are arranged in order of increasing variability on the right side, and their PCR success rates in the 16 tested vertebrates are shown on the left side. NPCLs indicated with asterisks may have extra copies in teleost genomes and thus are not suitable for phylogenetic studies of teleosts.

30 nuclear genes provide further support for the monophyly of lissamphibians and the Batrachia hypothesis (fig. 4). Additionally, all possible hypotheses against the monophyly of extant amphibians and the Batrachia hypothesis were rejected in our topological tests (table 2). However, the Batrachia hypothesis did not receive strong support in our species tree analysis ($BP_{MP\text{-}EST}$ = 74%; fig. 4), suggesting that more nuclear genes are still needed to resolve this node.

The monophyly of the internally fertilizing salamanders (Salamandroidea; all salamanders exclusive of Hynobiidae, Cryptobranchidae, and Sirenidae) is strongly supported in our analyses (fig. 4), in line with most recent molecular studies (Wiens et al. 2005; Roelants et al. 2007; Zhang and Wake 2009; Pyron and Wiens 2011) but differing strongly from Frost et al. (2006), who recovered a clade comprising Sirenidae, Dicamptodontidae, Ambystomatidae, and Salamandridae. The internally fertilizing salamanders include two well-supported clades: one is composed of Ambystomatidae, Dicamptodontidae, and Salamandridae, and the other of Proteidae, Rhyacotritonidae, Amphiumidae, and Plethodontidae (fig. 4).

Currently, two hypotheses have been proposed regarding the basal split within living salamanders. The traditional view favors Sirenidae as the sister group to all remaining salamanders (Duellman and Trueb 1994). This hypothesis received strong support in two recent studies (based on mitochondrial genomes, BP_{ML} = 98%; Zhang and Wake 2009; based on mitochondrial genomes and nuclear genes, $BP_{ML} > 80\%$; San Mauro 2010). In contrast, some studies suggest that the basal split separates Cryptobranchidae + Hynobiidae from all other salamanders (Gao and Shubin 2001; Wiens et al. 2005; Frost et al. 2006; Roelants et al. 2007; Pyron and Wiens 2011) but always without strong support ($BP_{ML} < 71\%$). Our phylogenetic analyses based on 30 independent NPCLs supported the second hypothesis that Cryptobranchoidea (Cryptobranchidae + Hynobiidae) branched first within the living salamanders. This result is extremely robust in our concatenation analyses (BP_{ML} = 99%, PP_{BAY} = 1.0, PP_{CAT} = 1.0; fig. 4) and statistically rejects all alternative hypotheses (table 2). In the species tree analysis without data concatenation, this result is also strong ($BP_{MP\text{-}EST}$ = 83%; fig. 4).

How many nuclear genes, then, are needed to robustly resolve the question of the basal split within living salamanders? Our analysis of data subsets indicates a progressive increase in the bootstrap support value for the node of interest (fig. 4) when an increasing number of genes are analyzed (fig. 5). Analyses based on single genes rarely resolve the node of interest with any confidence. Analyses based on 5–10 genes produce bootstrap support values of 60–80% in concatenation analyses (fig. 5), which is congruent with all previous nuclear studies using similar-sized data sets (Roelants et al. 2007; Pyron and Wiens 2011). Taking a bootstrap value of 95% in concatenation analyses as the threshold of "fully resolved," the minimum number of nuclear genes needed to resolve the root of the salamander tree is approximately 25. The previous contradictory results may be due to the overwhelmingly strong signals from the mitochondrial genome. Because

Table 1. Summary Information for the 30 NPCL Amplified in 19 Salamander Taxa.

Gene	Length (bp)	Taxa Amplified (%)	PCR Products Directly Sequenced (%)	GC%	Var. Sites (%)	PI Sites (%)	Overall Mean	
							P Distance	RCV
BPTF	552	19 (100)	19 (100)	43	163 (30)	118 (21)	0.098	0.093
CAND1	1,155	19 (100)	17 (89)	44	403 (35)	314 (27)	0.116	0.065
DET1	711	19 (100)	18 (95)	46	275 (39)	216 (30)	0.131	0.091
DISP1	960	19 (100)	19 (100)	41	317 (33)	211 (22)	0.096	0.072
DNAH3	918	19 (100)	19 (100)	42	389 (42)	304 (33)	0.139	0.049
DOLK	672	16 (84)	12 (75)	52	316 (47)	236 (35)	0.173	0.126
DSEL	1,266	19 (100)	19 (100)	44	546 (43)	415 (33)	0.148	0.055
ENC1	1,083	19 (100)	19 (100)	51	363 (34)	279 (26)	0.120	0.057
EXTL3	1,245	19 (100)	17 (89)	47	465 (37)	322 (26)	0.118	0.067
FAT4	738	19 (100)	19 (100)	45	344 (47)	249 (34)	0.156	0.072
FICD	510	18 (95)	18 (100)	44	169 (33)	124 (24)	0.111	0.074
GRM2	690	18 (95)	18 (100)	54	240 (35)	176 (26)	0.115	0.118
HYP	1,260	19 (100)	19 (100)	47	516 (41)	359 (28)	0.122	0.049
KBTBD2	1,059	19 (100)	19 (100)	44	406 (38)	246 (23)	0.103	0.040
KCNF1	735	19 (100)	19 (100)	52	294 (40)	220 (30)	0.151	0.153
KIAA1239	1,362	19 (100)	19 (100)	42	479 (35)	338 (25)	0.110	0.048
KIAA2013	516	19 (100)	19 (100)	52	221 (43)	178 (34)	0.148	0.093
LIG4	1,017	19 (100)	19 (100)	39	434 (43)	301 (30)	0.137	0.043
LPHN2	573	19 (100)	19 (100)	47	192 (34)	140 (24)	0.106	0.108
LRRN1	837	19 (100)	18 (95)	49	345 (41)	240 (29)	0.130	0.119
MGAT4C	747	18 (95)	18 (100)	42	273 (37)	212 (28)	0.126	0.079
MIOS	879	19 (100)	19 (100)	45	291 (33)	213 (24)	0.107	0.065
PANX2	717	19 (100)	19 (100)	44	254 (35)	199 (28)	0.125	0.059
PDP1	1,035	19 (100)	19 (100)	45	348 (34)	261 (25)	0.110	0.080
PPL	1,338	19 (100)	17 (89)	47	645 (48)	485 (36)	0.156	0.064
RAG1	1,380	19 (100)	19 (100)	51	550 (40)	438 (32)	0.146	0.072
RAG2	792	19 (100)	18 (95)	49	406 (51)	310 (39)	0.184	0.090
SACS	1,101	19 (100)	19 (100)	40	383 (35)	282 (26)	0.105	0.040
TTN	984	19 (100)	5 (26)	43	378 (38)	269 (27)	0.125	0.052
ZBED4	1,002	19 (100)	19 (100)	39	325 (32)	224 (22)	0.093	0.040

NOTE.—Length, length of refined alignment; Var. sites, variable sites; PI sites, parsimony informative sites; RCV, relative composition variability.

the initial diversification of salamanders occurred within a relatively short window of time (Weisrock et al. 2005), the genealogical histories of individual gene loci may sometimes appear misleading in terms of the relationships among species due to incomplete lineage sorting. Unfortunately, the mitochondrial genome recorded such an incorrect history.

Discussion

The NPCL toolkit and experimental protocol introduced here is a highly reliable, rapid, and cost-effective method for building medium-scale multilocus data to produce high-resolution phylogenetic relationships. This phylogenomic approach has the potential to accelerate the completion of many parts of the vertebrate tree of life because no further marker development is required, which is often the bottlenecks in phylogenetic research. Once a specific phylogenetic question within vertebrates arises, researchers simply need to check the list for our toolkit and look for NPCL markers with expected evolutionary rates and experimental performance for their groups of interest. Then many orthologous loci can be quickly obtained by traditional PCR and Sanger sequencing, usually without time-consuming gel cutting and cloning. Applying the NPCL toolkit may also have another benefit for assembling the vertebrate tree of life because people working on different groups can easily use the same set of loci, which will facilitate combined analyses.

Merits of the Toolkit

Because of the use of the nested PCR strategy outlined earlier, most NPCL in the toolkit work for all major jawed vertebrate groups with high experimental success rates (normally > 95%). Such results were achieved in unified PCR conditions without any extra effort involving cycling condition optimization. This feature of the toolkit enables it to surpass previously developed nuclear marker sets (Murphy et al. 2001; Li et al. 2007; Thomson et al. 2008; Townsend et al. 2008; Wright et al. 2008; Portik et al. 2011; Shen et al. 2011; Zhou et al. 2011). Most previous nuclear marker sets were developed for specific animal groups, and their application to

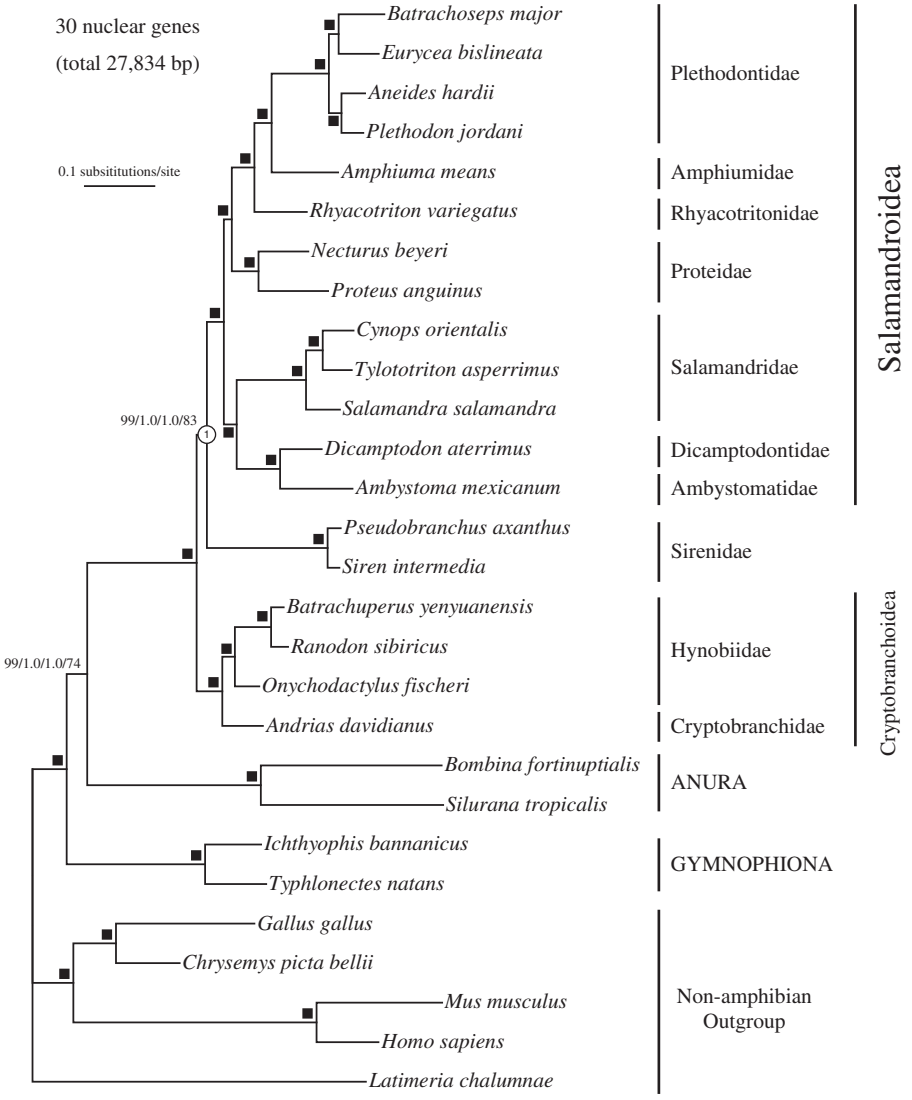

FIG. 4. Higher-level phylogenetic relationships of 10 salamander families inferred from 30 NPCL markers. The tree was inferred by concatenation analyses using ML, BI, and the mixture model (CAT) and by species-tree analysis using the pseudo-ML approach (MP-EST). Branch support values are indicated beside nodes in order of ML bootstrap (BP_{ML}), BI posterior probability (PP_{BI}), CAT posterior probability (PP_{CAT}), and MP-EST bootstrap (BP_{MP-EST}), from left to right. The filled squares represent $BP_{ML} > 95$, $PP_{BAY} = 1.0$, $PP_{CAT} = 1.0$, and $BP_{MP-EST} > 95$. The circled number refers to the node of interest studied in figure 6. Branch lengths are from the ML analysis.

Table 2. Statistical Confidence (*P* Values) for Alternative Branching Hypotheses Based on 30-Gene Data Set.

Alternative Topology Tested	ΔLn *L*	*P* Value			Rejection
		AU	BP	KH	
Best ML	0	0.993	0.97	0.98	– – –
Sirenidae branched earlier	32.8	0.025	0.015	0.02	+ + +
Sirenidae is sister to Cryptobranchoidea	42.3	0.004	0.002	0.004	+ + +
Gymnophiona is sister to Anura (monophyletic lissamphibians)	34.3	0.013	0.012	0.013	+ + +
Gymnophiona is sister to Caudata (monophyletic lissamphibians)	43.9	0.002	0.001	0.001	+ + +
Anura is sister to Amniota (paraphyletic lissamphibians)	144.6	5E−30	0	0	+ + +
Gymnophiona is sister to Amniota (paraphyletic lissamphibians)	129.0	1E−69	0	0	+ + +
Caudata is sister to Amniota (paraphyletic lissamphibians)	172.8	0.0001	0	0	+ + +

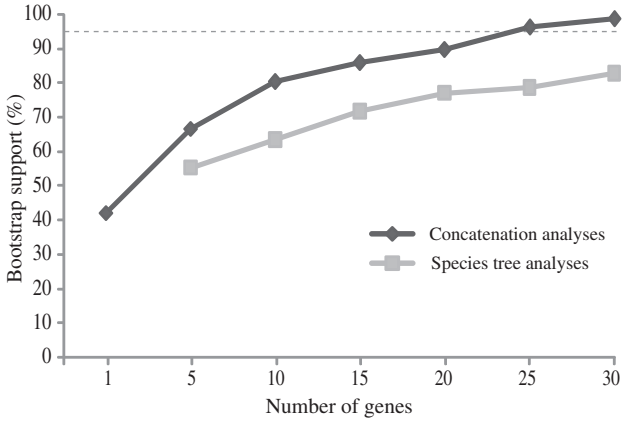

Fig. 5. The effect of increasing the number of nuclear loci on resolving the basal split within salamanders. Each data point represents the mean of support values estimated from 30 randomly sampled subsets. The dashed line indicates the threshold of 95% bootstrap support values. The statistical plots show that the minimum number of nuclear loci needed to robustly resolve the basal split within salamanders is 25.

other distantly related groups is usually difficult. For example, Spinks et al. (2010) collected 120 nuclear markers from avian, squamate, and mammalian phylogenetic studies and evaluated their PCR performance in turtles. They found that only eight nuclear markers successfully produced single, expected bands across 13 tested turtle species. In another case, Fong and Fujita (2011) developed 75 nuclear markers for vertebrate phylogenetics, but approximately 60% of the target fragments were unable to obtain in three test species (two reptiles and one lissamphibian). Therefore, although the nested PCR method introduced here requires an additional PCR reaction, the extra work is still worthwhile.

In PCR-based phylogenetic projects, even when the PCR reactions are successful, the products often contain significant nonspecific amplicons. Such a condition requires additional effort involving gel purification and cloning, which involves much more time than the PCR reaction. Our NPCL toolkit is specifically designed to solve this problem, so that normally, over 90% of PCR reactions produce strong and single expected bands. Moreover, most of the primers used to date for nuclear marker sets are degenerate and thus are not suitable for direct sequencing PCR products. Benefiting from the use of our nested PCR strategy, we introduce anchoring sequences to the ends of PCR fragments while maintaining PCR efficiency. Such introduced anchoring sequences bring the added benefit that two specific sequencing primers (Seq_F and Seq_R) can be used in all Sanger sequencing reactions.

One additional feature of our NPCL toolkit is that the average length of the NPCLs within it is 1,050 bp, a length that can easily be amplified in one PCR reaction and sequenced in both directions to allow efficient use of resources. In contrast, the average marker lengths are 920 bp for 10 NPCLs in Li et al. (2007), 760 bp for 26 NPCLs in Townsend et al. (2008), 873 bp for 22 NPCLs in Shen et al. (2011), and 470 bp for 75 NPCLs in Fong and Fujita (2011), respectively. Longer markers will provide more sites than shorter ones for equivalent money and time. This feature makes our NPCL toolkit more cost-effective than previously developed nuclear marker sets.

Phylogenetic Utility

The vertebrate NPCL toolkit we developed here shows great promise in terms of phylogenetic utility. A remarkable feature of our NPCL toolkit is that it provided 102 NPCLs with a broad range of evolutionary rates. In the case of our demonstration, we used 30 NPCLs to resolve a family-level salamander phylogeny using both traditional concatenation analyses and a more promising species-tree analysis. However, this example does not mean that our toolkit performs well only on deep-timescale questions. Our ongoing study using this toolkit to resolve the intra-relationships within Plethodontidae, a rapidly radiating group of salamanders, suggests that the toolkit developed here also performs well in resolving species-level phylogenies. For many vertebrate groups in which applicable nuclear markers are limited, such as some teleosts, frogs and salamanders, our NPCL toolkit can provide a one-stop solution for phylogenetic studies from the family level to the species level. Even for those groups in which specific nuclear marker sets have been developed, our toolkit is still worth trying, as many more loci can be easily obtained that may resolve some difficult branches.

The Toolkit Is a Good Addition to Sequence Capture Approaches

Recently, sequence capture approaches have been applied to vertebrate phylogenomics (Crawford et al. 2012; Faircloth et al. 2012; Lemmon et al. 2012; McCormack et al. 2012). These approaches begin with the selective capture of genomic regions. Briefly, fragmented gDNA is hybridized to DNA or RNA probes either on an array or in solution. Nontargeted DNA is then washed away, and the targeted DNA is sequenced through NGS. The most promising feature of the sequence capture approach is that it can simultaneously produce hundreds to thousands of loci for tens of individuals within a relatively short time. Therefore, the sequence capture approach is considered to be much more cost-effective than the PCR-based method. According to the calculation of Lemmon et al. (2012), for a 100 taxa × 500 loci project, the cost of the sequence capture method is just 1–3.5% of the PCR-based method.

However, the sequence capture approach is currently too challenging for most phylogenetic researchers. Typical NGS runs (454 or Illumina) used by the sequence capture method generate 1,000,000–2,000,000,000 sequences. Storing and processing these NGS data require significant computer memory, hardware upgrades, and bioinformatic programming skills, which are often not familiar to most phylogenetic researchers. Moreover, phylogenetic reconstruction assumes that orthologous genes are being analyzed across species. For the PCR-based method, the detection of paralogous genes is relatively straightforward. However, in the sequence capture method, the captured genomic regions comprise short conserved cores (probe regions) and long unconserved flanking

sequences. Because paralogy cannot be detected until after the data are aligned, those unalignable sequences will make the detection of paralogy more difficult.

In fact, not every phylogenetic project will use more than 500 loci as the sequence capture method normally does. Based on both empirical and simulation data, 20–50 loci are generally sufficient to answer many phylogenetic questions (Rokas et al. 2003; Spinks et al. 2009). This is also the number of loci that most phylogenetic studies will use. In such a situation, adopting the sequence capture method is not cost-effective because researchers need to use relatively expensive NGS sequencing and spend time learning new experimental techniques and carrying out sophisticated bioinformatic processing. Our NPCL toolkit is specially designed for such medium-scale phylogenetic projects using approximately 50 loci. Such a number of expected loci can be easily fulfilled with our 102 NPCLs. Because more than 90% of the PCR reactions generated by our toolkit can be directly sequenced, the average cost for one locus per sample is rather low. In our laboratory, generating one new sequence typically costs US$ 3 (without considering labor).

In addition, researchers sometimes have only tiny amounts of DNA, but they wish to perform a multilocus phylogenetic analysis. In such a situation, the sequence capture method is difficult to implement because it normally requires DNA at the microgram level (Lemmon et al. 2012). Our NPCL toolkit can fill the gap here. Benefiting from the use of the nested PCR strategy, the sensitivity of PCR reactions in our method is extremely high. In many test experiments in our laboratory, the toolkit and protocol could produce target bands with only 5–10 ng of DNA.

Our NPCL toolkit is an alternative to the sequence capture method for the everyday work of phylogenetic researchers. Which method to choose depends on two major drivers: the amounts of DNA and the expected number of loci. When your DNA is limited, the better solution may be PCR; otherwise, sequence capture also works. Taking into account the money and time the two methods require, we speculate that the economic transition point from PCR to sequence capture is at approximately 100 loci. That assessment is why our toolkit includes 102 NPCL markers. Our proposal is that when using $\leq$100 loci, one can try our NPCL toolkit; when using $>$100 loci, sequence capture should be used.

Future Directions

In this study, we used multiple genome alignments deposited in the University of California–San Cruz (UCSC) genome browser to identify long and conserved exons across jawed vertebrates. Benefiting from the use of a nested PCR strategy, the experimental performance of the developed NPCLs indicated that they are highly stable in all major jawed vertebrate groups. Recently, a database for mining exon and intron markers, called EvolMarkers, has been built by Li et al. (2012). Careful investigation of this database may identify many conserved exons within nonvertebrates, whose interrelationships are currently more problematic than those of vertebrates. Because the nonvertebrates constitute many distantly related groups, it may be impossible to develop a single set of PCR primers for all nonvertebrates. However, following a similar marker development strategy, multiple NPCL toolkits could be constructed for various groups of nonvertebrates such as arthropods, echinoderms, and molluscs. In addition, because introns are flanked by conserved exons, the idea of the use of nested PCRs for marker development could also be applied to the development of EPIC (exon-primed intron crossing) markers, which are more suitable in shallow-scale phylogenetic or phylogeographic projects.

Despite the benefits of our proposed method, it must be recognized that when handling large-scale projects such as 200 taxa $\times$ 100 loci, the use of our toolkit and Sanger sequencing will still require significant cost, time, and labor. An alternative solution is to use NGS to replace Sanger sequencing. Recently, 454 NGS technology has been applied to sequence-targeted gene regions from a pool of PCR products from different specimens (Binladen et al. 2007; Meyer et al. 2008). In such experiments, specific tagging sequences must be added to amplicons by either PCR (Binladen et al. 2007) or blunt-end ligation (Meyer et al. 2008). Therefore, if the tailing sequences of the second-round PCR primers in our NPCL toolkit are replaced by tagging sequences instead (for tag designing, see Faircloth and Glenn 2012), all PCR products can be pooled together and sequenced with the 454 NGS, which will greatly reduce the money and time cost compared with Sanger sequencing. However, parallel tagged sequencing via NGS does not circumvent the process of PCR for each individual at each locus, which may be the most onerous part of a large-scale phylogenomic project. Some promising new technologies may help to solve this problem, such as microdroplet PCR (Tewhey et al. 2009), where millions of individual PCR reactions are performed in picoliter-scale droplets simultaneously, and the 96.96 Dynamic Array by Fluidigm, which allows 96 primer combinations to be used on 96 samples (9,216 total PCR reactions) on a single PCR plate. However, there has been little research to applying NGS and new high-throughput PCR technologies to phylogenomics, so their ease-of-use and cost-effectiveness still need to be explored.

Summary

In conclusion, we have developed an improved method for rapidly amplifying and sequencing NPCLs that has proven to be useful and effective for molecular phylogenetic studies of vertebrates. The newly developed toolkit provides an attractive alternative to available methods for vertebrate phylogenomics.

Materials and Methods

Development of NPCL and Primer Design

Our previous study showed that nested PCR is overwhelmingly more effective than conventional PCR for obtaining target amplicons from complex genomic environments (Shen et al. 2012). However, nested PCR requires four conserved regions to design two pairs of primers (illustrated in fig. 6, yellow blocks represent the conserved regions used for primer design), which means that only relatively long exons are

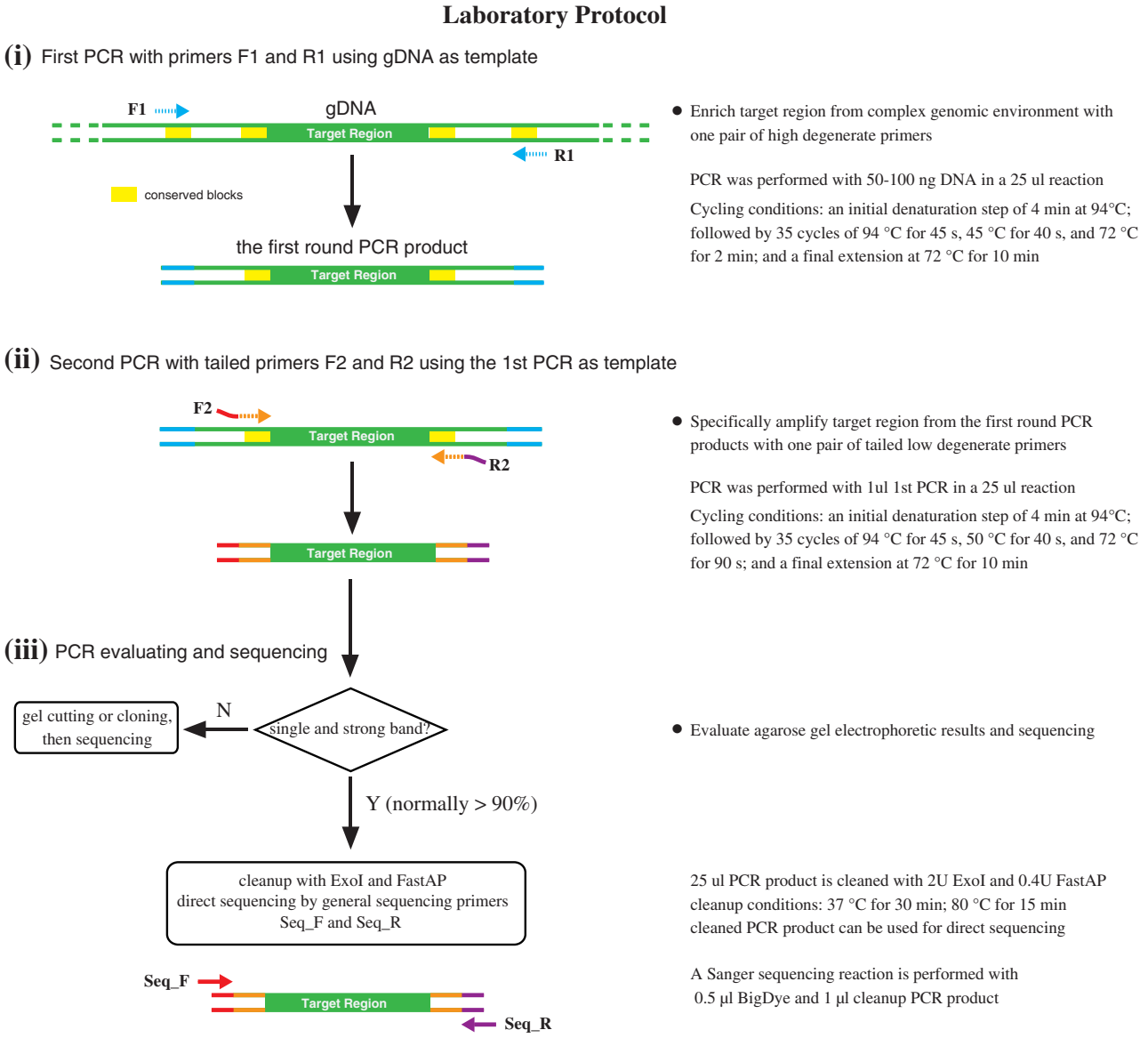

FIG. 6. Schematic representation of the experimental protocol for using our NPCL toolkit. Note that for each NPCL, nested PCR primers are designed on four short conserved blocks flanking the target region.

suitable as candidates for NPCL development with the nested PCR method. To search for long and conserved exons, we took advantages of our previous bioinformatic method, which used the multiple genome alignment data from the UCSC Genome Browser to identify conserved exons (Shen et al. 2011). Because the NPCL markers are to be used in vertebrates, we focused only on those multiple genome alignments that include at least six species: *Danio rerio* (zebrafish), *Silurana tropicalis* (frog), *Anolis carolinensis* (lizard), *Gallus gallus* (chicken), *Mus musculus* (mouse), and *Homo sapiens* (human). The alignments of candidate exons had to meet two criteria: length of more than 700 bp and pairwise similarity ranging from 35% to 90%. The detailed bioinformatic pipeline has been described elsewhere (Shen et al. 2011). In addition to using multiple genome alignments to screen NPCL candidates, we also manually searched for nuclear genes that were used previously (Murphy et al. 2001; Li et al. 2007; Townsend et al. 2008; Wright et al. 2008; Zhou et al. 2011; Song et al. 2012) in the ENSEMBL database to check whether they contain large and appropriately conserved exons.

As a result, we assembled a total of 305 NPCL candidate alignments, of which 120 contained the appropriate number of conserved blocks, and used these to design nested PCR primers. To increase the success rates of our NPCL markers in amniotes, we manually added a turtle sequence (*Chrysemys picta bellii*) to each of the candidate alignments using data downloaded from the ENSEMBL database. A total number of 480 primers were designed for the 120 NPCL candidates. Briefly, the first-round PCR primers are only used to enrich target regions from genomic environments and not to obtain target amplicons, so the degeneracy of these primers is normally high to increase reaction sensitivity; the second-round PCR primers are used to obtain target amplicons, so the

degeneracy of these primers is lower to increase reaction specificity. Our previous study showed that the nested PCR method often produces strong and single amplicon bands (Shen et al. 2012). To facilitate the next-step direct sequencing, we added a tail (5′-AGGGTTTTCCCAGTCACGAC-3′) to the 5′-end of all second-round forward primers and a tail (5′-AGATAACAATTTCACACAGG-3′) to the 5′-end of all second-round reverse primers. These tail sequences can provide two unique anchoring sites for direct sequencing from cleaned PCR products. In our pilot experiments, adding the tail sequences to primers did not affect the efficiency of the second-round PCR.

Experimental Testing for Candidate Markers in 16 Jawed Vertebrates

To test the experimental performance of our newly designed NPCL markers, we selected 16 taxa representing nine major jawed vertebrate lineages: Chondrichthyes (*Sphyrna lewini*); Actinopterygii (*Lepisosteus oculatus* and *Pangasius sutchi*); Dipnoi (*Protopterus annectens*); Lissamphibia (*Ichthyophis bannanicus*, *Batrachuperus yenyuanensis*, and *Rana nigromaculata*); Mammalia (*Mus musculus* and *Sus scrofa domestica*); Testudines (*Trionyx sinensis* and *Podocnemis unifilis*); Aves (*Struthio camelus* and *Zosterops japonica*); Crocodylia (*Crocodylus siamensis*); and Squamata (*Hemidactylus bowringii* and *Naja naja atra*). Total genomic DNA was extracted from ethanol-preserved tissues (liver or muscle) using the standard salt extraction protocol. All extracted genomic DNAs were diluted to a concentration of 50 ng μl^{-1} with 1 × TE and stored at −20 °C before PCR amplification.

All 120 NPCL markers were tested with a two-round PCR strategy (nested PCR). The first-round PCR was performed in 25 μl reaction containing 1–2 μl template DNA (50–100 ng), with final concentrations of 1× PCR buffer, 200 μM dNTP, 400 nM of each forward and reverse first-round primers, and 1.25 U Taq polymerase (TransTaq High Fidelity; TransGen, Beijing). The cycling conditions of the first-round PCR were as follows: an initial denaturation step of 4 min at 94 °C followed by 35 cycles of a 45 s denaturation at 94 °C, a 40 s annealing at 45 °C, and a 2 min extension at 72 °C followed by a final 10 min extension at 72 °C. The second-round PCR was also performed in 25 μl reaction containing 1 μl of the first round PCR product (without dilution) and final concentrations of 1× PCR buffer, 200 μM dNTP, 400 nM of each forward and reverse second-round primers, and 1.25 U Taq polymerase. The cycling conditions of the second-round PCR were as follows: an initial denaturation step of 4 min at 94 °C followed by 35 cycles of a 45 s denaturation at 94 °C, a 40 s annealing at 50 °C, and a 90 s extension at 72 °C followed by a final 10 min extension at 72 °C.

One microliter of the second-round PCR products was analyzed on 1.0% TAE agarose gel. An NPCL marker was considered successful if more than 8 of the 16 tested taxa produced target amplicon bands. On this basis, 102 out of 120 tested NPCL markers were successful. The nested-PCR primers for the 102 NPCL markers can be found in the online supplementary table S1, Supplementary Material online. If the PCR products contained significant nonspecific amplicon bands (normally < 10%), they needed further processing, for example, standard gel cutting or cloning. If the PCR reactions produced single amplicon bands (normally > 90%), they were cleaned with ExoFAP treatment: 2 U ExoI and 0.4 U FastAP (all Fermentas) were added to the PCR tube and incubated for 30 min at 37 °C and 15 min at 80 °C. The cleanup PCR reactions can be directly used as templates for Sanger sequencing. According to our experimental designs, all PCR fragments can be sequenced with the two universal sequencing primers Seq_F: 5′-AGGGTTTTCCCAGTCACGAC-3′ and Seq_R: 5′-AGATAACAATTTCACACAGG-3′ from both ends. A typical Sanger sequencing reaction in our laboratory consumes 0.5 μl BigDye and 1 μl cleanup PCR product. The primer design strategy, the laboratory protocol for the nested PCR method and the pretreatment of PCR products before Sanger sequencing are illustrated in figure 6.

Calculation of Relative Evolutionary Rate of 102 NPCLs

The rate multipliers (m) across partitions estimated in MrBayes 3.2 (Ronquist et al. 2012) are used as relative evolutionary rates. To calculate these parameters, alignments for each NPCL were prepared for 12 species: *Homo sapiens*, *Macaca mulatta*, *Mus musculus*, *Rattus norvegicus*, *Gallus gallus*, *Meleagris gallopavo*, *Chrysemys picta bellii*, *Anolis carolinensis*, *Silurana tropicalis*, *Tetraodon nigroviridis*, *Takifugu rubripes*, and *Danio rerio*. Because genome data are available for the 12 species, we did not generate any new data. The 102 NPCL alignments were then combined and subjected to MrBayes analyses, partitioned by genes. Each gene was assigned a separate GTR + Γ + I model and all model parameters were unlinked. Two Markov chain Monte Carlo (MCMC) runs were performed with one cold chain and three heated chains (temperature set to 0.1) for 50 million generations and sampled every 1,000 generations. The rate multiplier for each gene was estimated using Tracer version 1.4 after discarding the first 50% of the generations. All evolutionary rates were normalized by dividing by the maximum value of the obtained rates.

Gene and Taxon Sampling for Investigating Higher Level Salamander Relationships

To test the utility of our NPCL toolkit in a real case, we selected 19 salamander taxa representing all 10 salamander families and 9 outgroup taxa to investigate the family-level relationships of salamanders (supplementary table S2, Supplementary Material online). For gene sampling, we randomly selected 30 NPCL markers whose PCR success rates were more than 90% in the 16 previously tested vertebrates. Among the target 840 sequences (30 markers for 28 taxa), 201 were available in public databases (NCBI, UCSC, and ENSEMBL), whereas the remaining 639 sequences needed to be generated de novo. The experimental procedure was as described earlier. All obtained sequences were examined by checking for the presence of premature stop codons (pseudogene) and by BlastX searches against the nonredundant

protein sequences (nr) to confirm that they were our target genes. The NPCLs, species, and accession numbers for the newly obtained sequences are listed in the supplementary table S2, Supplementary Material online.

Phylogenetic Analyses

Alignments of all 30 NPCL markers were conducted using the G-INS-i method from MAFFT (Katoh et al. 2005) under the default settings according to their translated amino acid sequences, then refined by eye. All 30 refined alignments were combined into a concatenated data set (27,834 bp).

For the concatenated data set, we manually defined five partitioning strategies: 2 partitions (one for codon positions 1 and 2 and one for codon position 3); 3 partitions (one partition for each codon position); 30 partitions (one partition for each gene); 60 partitions (one for codon position 1 + 2 and one codon position 3 across 30 genes); and 90 partitions (codon position partitioning across 30 genes). Comparisons of the five partitioning strategies and selections of corresponding nucleotide substitution models were conducted under the Bayesian information criterion implemented in PartitionFinder (Lanfear et al. 2012). The 3-partition scheme (one partition for each codon position) was chosen as the best-fitting partitioning strategy, and all 3 partitions favored the GTR + Γ + I model.

The concatenated data set was separately analyzed with both ML and Bayesian inference (BI) methods under the 3-partition scheme. Partitioned ML analyses were implemented using RAxML version 7.2.6 (Stamatakis 2006), with the GTR + Γ + I model assigned for each partition. A search that combined 100 separate ML searches was applied to find the optimal tree, and branch support for each node was evaluated with 500 standard bootstrapping replicates (-f d -b 500 option) implemented in RAxML. The partitioned BI was conducted using MrBayes 3.2 (Ronquist et al. 2012). All model parameters were unlinked. Two MCMC runs were performed with one cold chain and three heated chains (temperature set to 0.1) for 60 million generations and sampled every 1,000 generations. The chain stationarity was visualized by plotting −ln *L* against the generation number using Tracer version 1.4, and the first 50% of generations were discarded. Topologies and posterior probabilities were estimated from the remaining generations. Two runs for each analysis were compared for congruence.

We also performed Bayesian phylogenetic analyses under a mixture model CAT + GTR + Γ4 in PhyloBayes 3.3 (Lartillot et al. 2009) with two independent MCMC runs. Each run was performed for 10,000 cycles and sampled every cycle. Stationarity was reached when the largest discrepancy (maxdiff) was less than 0.1 between two independent runs. The first 5,000 trees in each MCMC run were discarded. The remaining 10,000 trees of the two runs were sampled every 5 trees to generate a 50% majority-rule posterior consensus tree.

Species tree estimation was conducted using the pseudo-ML approach in the program MP-EST (Liu et al. 2010) under the coalescent model. Briefly, the gene trees for 30 NPCL markers were reconstructed with ML under the GTR + Γ8 model using PHYML 3.0 (Guindon et al. 2010). The resulting 30 gene trees were rooted with an outgroup (*Chrysemys picta bellii*) and used to generate an MP-EST species tree using the program MP-EST. The robustness of the species tree was evaluated with nonparametric bootstrapping of 500 replicates.

Alternative phylogenetic hypotheses were tested based on 30-gene data set. We first calculated sitewise log likelihoods of alternative trees using RAxML (-f g) under the GTR + Γ + I model. Then, the site log-likelihood file was used to estimate *P* values for each alternative tree in CONSEL program (Shimodaira and Hasegawa 2001), using the Kishino–Hasegawa test (KH) (Kishino and Hasegawa 1989), the approximately unbiased test (AU) (Shimodaira 2002), and the RELL bootstrap proportion test (BP).

Supplementary Material

Supplementary tables S1 and S2 and figures S1 and S2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors are grateful to David Wake and Carol Spencer of the Museum of Vertebrate Zoology at the University of California, Berkeley for tissue loans. David Wake provided many useful comments on the manuscript. This work was supported by National Natural Science Foundation of China grants (31172075 and 30900136) and the National Science Fund for Excellent Young Scholars (no. pending).

References

Binladen J, Gilbert MTP, Bollback JP, Panitz F, Bendixen C, Nielsen R, Willerslev E. 2007. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS One* 2:e197.

Crawford NG, Faircloth BC, McCormack JE, Brumfield RT, Winker K, Glenn TC. 2012. More than 1000 ultraconserved elements provide evidence that turtles are the sister group to archosaurs. *Biol Lett.* 8: 783–786.

Delsuc F, Brinkmann H, Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. *Nat Rev Genet.* 6:361–375.

Duellman WE, Trueb L. 1994. Biology of amphibians. Baltimore (MD): Johns Hopkins University Press.

Dunn CW, Hejnol A, Matus DQ, et al. (18 co-authors). 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452:745–749.

Faircloth BC, Glenn TC. 2012. Not all sequence tags are created equal: designing and validating sequence identification tags robust to indels. *PLoS One* 7:e42543.

Faircloth BC, McCormack JE, Crawford NG, Brumfield RT, Glenn TC. 2012. Ultraconserved elements anchor thousands of genetic markers for target enrichment spanning multiple evolutionary timescales. *Syst Biol.* 61:717–726.

Fong JJ, Brown JM, Fujita MK, Boussau B. 2012. A phylogenomic approach to vertebrate phylogeny supports a turtle-archosaur affinity and a possible paraphyletic lissamphibia. *PLoS One* 7: e48990.

Fong JJ, Fujita MK. 2011. Evaluating phylogenetic informativeness and data-type usage for new protein-coding genes across Vertebrata. *Mol Phylogenet Evol.* 61:300–307.

Frost DR, Grant T, Faivovich J, et al. (18 co-authors). 2006. The amphibian tree of life. *Bull Am Mus Nat Hist.* 297:8–370.

Gao KQ, Shubin NH. 2001. Late Jurassic salamanders from northern China. *Nature* 410:574–577.
Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59:307–321.
Hugall AF, Foster R, Lee MSY. 2007. Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. *Syst Biol.* 56:543–563.
Inoue JG, Miya M, Lam K, Tay BH, Danks JA, Bell J, Walker TI, Venkatesh B. 2010. Evolutionary origin and phylogeny of the modern holocephalans (Chondrichthyes: Chimaeriformes): a mitogenomic perspective. *Mol Biol Evol.* 27:2576–2586.
Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 33:511–518.
Kishino H, Hasegawa M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J Mol Evol.* 29: 170–179.
Künstner A, Wolf JBW, Backström N, et al. (13 co-authors). 2010. Comparative genomics based on massive parallel transcriptome sequencing reveals patterns of substitution and selection across 10 bird species. *Mol Ecol.* 19:266–276.
Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol.* 29:1695–1701.
Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25:2286–2288.
Lemmon AR, Brown JM, Stanger-Hall K, Lemmon EM. 2009. The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. *Syst Biol.* 58:130–145.
Lemmon AR, Emme SA, Lemmon EM. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Syst Biol.* 61:727–744.
Li C, Ortí G, Zhang G, Lu G. 2007. A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. *BMC Evol Biol.* 7:44.
Li C, Riethoven JJ, Naylor GJ. 2012. EvolMarkers: a database for mining exon and intron markers for evolution, ecology and conservation studies. *Mol Ecol Resour.* 12:967–971.
Liu L, Yu L, Edwards SV. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evol Biol.* 10:302.
McCormack JE, Faircloth BC, Crawford NG, Gowaty PA, Brumfield RT, Glenn TC. 2012. Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. *Genome Res.* 22: 746–754.
McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. *Mol Phylogenet Evol.* 66:526–538.
Meyer A, Van de Peer Y. 2005. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *Bioessays* 27:937–945.
Meyer M, Stenzel U, Hofreiter M. 2008. Parallel tagged sequencing on the 454 platform. *Nat Protoc.* 3:267–278.
Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ. 2001. Molecular phylogenetics and the origins of placental mammals. *Nature* 409:614–618.
Philippe H, Derelle R, Lopez P, et al. (20 co-authors). 2009. Phylogenomics revives traditional views on deep animal relationships. *Curr Biol.* 19:706–712.
Philippe H, Telford M. 2006. Large-scale sequencing and the new animal phylogeny. *Trends Ecol Evol.* 21:614–620.
Portik DM, Wood PL, Grismer JL, Stanley EL, Jackman TR. 2011. Identification of 104 rapidly-evolving nuclear protein-coding markers for amplification across scaled reptiles using genomic resources. *Conserv Genet Resour.* 4:1–10.
Pyron RA, Wiens JJ. 2011. A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. *Mol Phylogenet Evol.* 61: 543–583.
Roelants K, Gower DJ, Wilkinson M, Loader SP, Biju SD, Guillaume K, Moriau L, Bossuyt F. 2007. Global patterns of diversification in the history of modern amphibians. *Proc Natl Acad Sci U S A.* 104: 887–892.
Rokas A, Williams BL, King N, Carroll SB. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425:798–804.
Ronquist F, Teslenko M, Van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 61:539–542.
Roure B, Baurain D, Philippe H. 2013. Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. *Mol Biol Evol.* 30:197–214.
San Mauro D. 2010. A multilocus timescale for the origin of extant amphibians. *Mol Phylogenet Evol.* 56:554–561.
San Mauro D, Vences M, Alcobendas M, Zardoya R, Meyer A. 2005. Initial diversification of living amphibians predated the breakup of Pangaea. *Am Nat.* 165:590–599.
Shen XX, Liang D, Wen JZ, Zhang P. 2011. Multiple genome alignments facilitate development of NPCL markers: a case study of tetrapod phylogeny focusing on the position of turtles. *Mol Biol Evol.* 28: 3237–3252.
Shen XX, Liang D, Zhang P. 2012. The development of three long universal nuclear protein-coding locus markers and their application to osteichthyan phylogenetics with nested PCR. *PLoS One* 7: e39256.
Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. *Syst Biol.* 51:492–508.
Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17: 1246–1247.
Song S, Liu L, Edwards SV, Wu S. 2012. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proc Natl Acad Sci U S A.* 109: 14942–14947.
Spinks PQ, Thomson RC, Barley AJ, Newman CE, Shaffer HB. 2010. Testing avian, squamate, and mammalian nuclear markers for cross amplification in turtles. *Conserv Genet Resour.* 2:127–129.
Spinks PQ, Thomson RC, Lovely GA, Shaffer HB. 2009. Assessing what is needed to resolve a molecular phylogeny: simulations and empirical data from Emydid turtles. *BMC Evol Biol.* 9:56.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
Tewhey R, Warner JB, Nakano M, Libby B, Medkova M, David PH, Kotsopoulos SK, Samuels ML, Hutchison JB, Larson JW. 2009. Microdroplet-based PCR enrichment for large-scale targeted sequencing. *Nat Biotechnol.* 27:1025–1031.
Thomson RC, Shedlock AM, Edwards SV, Shaffer HB. 2008. Developing markers for multilocus phylogenetics in non-model organisms: A test case with turtles. *Mol Phylogenet Evol.* 49:514–525.
Thomson RC, Wang IJ, Johnson JR. 2010. Genome-enabled development of DNA markers for ecology, evolution and conservation. *Mol Ecol.* 19:2184–2195.
Townsend TM, Alegre RE, Kelley ST, Wiens JJ, Reeder TW. 2008. Rapid development of multiple nuclear loci for phylogenetic analysis using genomic resources: an example from squamate reptiles. *Mol Phylogenet Evol.* 47:129–142.
Weisrock DW, Harmon LJ, Larson A. 2005. Resolving deep phylogenetic relationships in salamanders: analyses of mitochondrial and nuclear genomic DNA. *Syst Biol.* 54:758–777.
Wiens J, Bonett R, Chippindale P. 2005. Ontogeny discombobulates phylogeny: paedomorphosis and higher-level salamander relationships. *Syst Biol.* 54:91–110.

Wright TF, Schirtzinger EE, Matsumoto T, et al. (11 co-authors). 2008. A multilocus molecular phylogeny of the parrots (Psittaciformes): support for a Gondwanan origin during the cretaceous. *Mol Biol Evol.* 25:2141–2156.

Zhang P, Wake DB. 2009. Higher-level salamander relationships and divergence dates inferred from complete mitochondrial genomes. *Mol Phylogenet Evol.* 53:492–508.

Zhang P, Zhou H, Chen YQ, Liu YF, Qu LH. 2005. Mitogenomic perspectives on the origin and phylogeny of living amphibians. *Syst Biol.* 54:391–400.

Zhou XM, Xu SX, Zhang P, Yang G. 2011. Developing a series of conservative anchor markers and their application to phylogenomics of Laurasiatherian mammals. *Mol Ecol Resour.* 11: 134–140.