Received: 9 June 2021 | Revised: 8 September 2021 | Accepted: 14 September 2021

DOI: 10.1111/1755-0998.13517

MOLECULAR ECOLOGY RESOURCES WILEY

RESOURCE ARTICLE

FrogCap: A modular sequence capture probe-set for phylogenomics and population genetics for all frogs, assessed across multiple phylogenetic scales

Carl R. Hutter[1] | Kerry A. Cobb[1] | Daniel M. Portik[2] | Scott L. Travers[1,3] | Perry L. Wood Jr.[1] | Rafe M. Brown[1]

[1]Biodiversity Institute and Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, Kansas, USA

[2]California Academy of Sciences, San Francisco, California, USA

[3]Department of Biological Sciences, Rutgers University-Newark, Newark, New Jersey, USA

Correspondence
Carl R. Hutter, Museum of Natural Sciences and Department of Biological Sciences, Louisiana State University, Baton Rouge, LA, USA.
Email: carl.hutter@gmail.com

Present address
Kerry A. Cobb and Perry L. Wood, Department of Biological Sciences & Museum of Natural History, Auburn University, Auburn, Alabama, USA

Funding information
National Science Foundation, Grant/Award Number: 0743491, 0907996, 1451148, 1540502, 1557053, 1654388 and 1701952; KU Research Investment Council, Grant/Award Number: 2300207; Foundation for the National Institutes of Health, Grant/Award Number: K12GM093854

Abstract

Despite the prevalence of high-throughput sequencing in phylogenetics, many relationships remain difficult to resolve because of conflicting signal among genomic regions. Selection of different types of molecular markers from different genomic regions is required to overcome these challenges. For evolutionary studies in frogs, we introduce the publicly available FrogCap suite of genomic resources, which is a large collection of ~15,000 markers that unifies previous genetic sequencing efforts. FrogCap is designed to be modular, such that subsets of markers and SNPs can be selected based on the desired phylogenetic scale. FrogCap uses a variety of marker types that include exons and introns, ultraconserved elements, and previously sequenced Sanger markers, which span up to 10,000 bp in alignment lengths; in addition, we demonstrate potential for SNP-based analyses. We tested FrogCap using 121 samples distributed across five phylogenetic scales, comparing probes designed using a consensus- or exemplar genome-based approach. Using the consensus design is more resilient to issues with sensitivity, specificity, and missing data than picking an exemplar genome sequence. We also tested the impact of different bait kit sizes (20,020 vs. 40,040) on depth of coverage and found triple the depth for the 20,020 bait kit. We observed sequence capture success (i.e., missing data, sequenced markers/bases, marker length, and informative sites) across phylogenetic scales. The incorporation of different marker types is effective for deep phylogenetic relationships and shallow population genetics studies. Having demonstrated FrogCap's utility and modularity, we conclude that these new resources are efficacious for high-throughput sequencing projects across variable timescales.

KEYWORDS
amphibians, anura, exon capture, genomics, target capture, UCEs

1 | INTRODUCTION

Although whole-genome sequencing for non-model organisms remains costly, reduced representation strategies can create large phylogenomic and population genetic data sets at reduced financial costs (Glenn, 2011; Kircher & Kelso, 2010; Rohland & Reich, 2012). These approaches aim to more affordably sequence subsets from across the genome, while allowing more specimens to be sampled simultaneously (Rohland & Reich, 2012). An important consideration for systematics and population genetics is balancing the trade-off between obtaining orthologous genetic markers (defined here as DNA sequences with a known genomic location enabling

comparisons among samples) across moderate to deep timescales, while obtaining markers that are variable enough to resolve ambiguous phylogenetic relationships and provide high depth of coverage variants for population genetics (Jones & Good, 2016; Sulonen et al., 2011).

There are now many methodologies for reduced representation sequencing, and the choice largely depends on the desired phylogenetic breadth and number of samples. The most common methods for obtaining subsets of genome-wide sequence data include: (1) restriction-site associated digestion methods (RADseq): targets areas adjacent to restriction enzyme sites (Miller et al., 2007); (2) targeted sequence capture: targets genomic regions through hybridization-based sequence capture (Choi et al., 2009; Hancock-Hanser et al., 2013); and (3) transcriptome sequencing (RNASeq): targets the expressed exome of a sampled tissue type (Wang et al., 2009). These three methods have been used to address a variety of phylogenetic and population genetic questions and have method-specific advantages and disadvantages related to phylogenetic scale and reusability (reviewed in Lemmon & Lemmon, 2013 and McCormack et al., 2013). Trade-offs include financial costs and availability of pre-existing genomic resources to identify genomic regions of interest.

Selection of different types of molecular data (e.g., protein-coding exons, non-coding regions, ultra-conserved elements [UCEs], or single nucleotide polymorphisms [SNPs]) is an increasingly important issue when dealing with difficult phylogenetic questions. Different marker types can demonstrate conflicting phylogenomic results, vary in their utility at different timescales, and present biases for analyses (Hosner et al., 2016; Lemmon & Lemmon, 2013; Liu et al., 2009; McCormack et al., 2013). For example, there is debate over whether protein-coding or non-coding markers are more appropriate for phylogenetic analyses, because selection on protein-coding exons may mask homology and bias phylogenetic signal (Chen et al., 2017; Edwards, 2009; Jarvis et al., 2014; Liu et al., 2009; Reddy et al., 2017). One well-studied case in birds exemplifies this debate, in which phylogenomic studies have produced nontrivial differences in phylogeny based on different marker types (McCormack et al., 2013; Reddy et al., 2017). In addition, other marker features such as locus length (Edwards et al., 2016; Springer & Gatesy, 2018) or character type (e.g., SNPs: Leaché & Oaks, 2017; indels: Simmons & Ochoterena, 2000; transposable elements: Han et al., 2011) have been explored in this context.

These debates have made it clear that there is no consensus regarding a perfect, universal molecular marker type. One solution is to use an assortment of markers sampled from across the genome with different properties (Chen et al., 2017; Dool et al., 2016). The inclusion of a diversity of marker types allows downstream filtering (e.g., using length or informativeness) which can improve phylogenetic estimates and diagnosing causes of phylogenetic uncertainty (Chakrabarty et al., 2017; Karin et al., 2020; Mirarab et al., 2014; Springer & Gatesy, 2018; Streicher et al., 2018). When considering these factors, a sequence capture approach that targets a variety of marker types may represent an optimal solution. However, substantial genomic resources are required to develop the initial target panel and corresponding probe design for targeted sequence capture.

Targeted sequence capture of exons and other markers are widely used by the phylogenetics community. It has been largely dominated by targeting conserved elements using two main approaches: anchored hybrid enrichment for conserved exons (AHE, Lemmon et al., 2012) and ultra-conserved elements (UCEs, Faircloth et al., 2012). Both approaches identify regions of the genome that are conserved across distantly related taxa and use probes designed from these conserved regions for target capture (Gnirke et al., 2009). The two methods differ in that AHE predominantly targets several hundred medium length exons (>500 bp) that are moderately divergent but sufficiently conserved to capture regions across hundreds of millions of years (Lemmon et al., 2012). In contrast, the UCE approach targets several thousand ultra-conserved regions (~120 bp long) with the goal of obtaining variable flanking regions adjacent to the conserved regions (Bejerano et al., 2004; Faircloth et al., 2012). Both approaches have limitations, for example, AHE markers have the potential to bias phylogenetic results due to directional, divergent, or convergent selection (AHE targets only exons, Bragg et al., 2016; Castoe et al., 2009; Singhal et al., 2017). Regarding UCEs, their function is often unknown in vertebrates (Alexander et al., 2010), and they may be undergoing purifying selection (Katzman et al., 2007). Additionally, the variable flanking regions of UCEs can be difficult to align across distantly related taxa (Portik & Wiens, 2021; Singhal et al., 2017; Streicher et al., 2018). Both AHE and UCEs are widely used in inferring phylogenies across broad and shallow phylogenetic scales (Brandley et al., 2015; Crawford et al., 2012; Prum et al., 2015; Smith et al., 2014; Streicher et al., 2018). For population genetics, UCEs have been used with some success (Andermann et al., 2019; Hosner et al., 2016; Zarza et al., 2016). In contrast, AHE produces a smaller number of markers (<400, with some duplication issues; Karin et al., 2020) and generally results in relatively fewer independent SNPs as required by many types of analyses (Hedge & Wilson, 2014; Lanier & Knowles, 2012; Springer & Gatesy, 2018).

Amphibians represent one of the most diverse terrestrial vertebrate groups, and most extant diversity is concentrated within a single lineage – Anura (frogs and toads; called frogs hereafter). Frogs have been diversifying for over 200 million years and now include over 7,000 described species (88% of amphibians; AmphibiaWeb, 2021). This deep time scale and exceptional diversity has presented a major challenge for research into amphibian evolution. The performance of a universal frog custom capture kit remains largely unexplored (but see Hedtke et al., 2013; Portik et al., 2016; Salamanders: McCartney-Melstad et al., 2016). Several studies have used UCEs (Alexander et al., 2017; Pie et al., 2018; Streicher et al., 2018; Streicher et al., 2020; Zarza et al., 2016; Guillory et al., 2019; Guillory et al., 2020; Barrientos et al., 2021) and others have used AHE markers (Heinicke et al., 2018; Hime et al., 2021; Peloso et al., 2016; Yuan et al., 2019), whereas two studies created customized transcriptome-based probe-sets for an African frog clade (Afrobatrachia) and the Asian genus *Limnonectes* (Portik et al., 2016;

Reilly et al., 2019). Although UCEs have been used in frogs, they are not ideal because they were designed for amniotes (the sister clade to amphibians) therefore only about half the target UCEs are captured (~2500/5600 UCEs; Streicher et al., 2018; Guillory et al., 2019). The AHE probe-set is advantageous because it produces long exons where sequence evolution can be modelled more predictably. However, longer exons require more probes for suitable capture efficiency, and this reduces the number of markers that can be targeted with the same number of probes (Bragg et al., 2016; Singhal et al., 2017). In summary, despite the increased application of high-throughput sequencing in frogs, a universal probe-set incorporating different molecular marker types has not been developed specifically for frogs.

We present FrogCap, a publicly available collection of molecular markers for all frogs that can be adapted into sequence capture probe-sets. Importantly, it includes markers from a variety of data types (exons, introns, UCEs, and independent markers). We developed and tested a large set of markers and probes which correspond to ~15,000 markers. FrogCap unifies new markers with previous sequencing via the inclusion of "legacy" Sanger sequencing markers traditionally used in phylogenetic studies (Table 1; Feng et al., 2017; Frost et al., 2006; Pyron & Wiens, 2011) and UCEs that have been successfully captured in anurans (Alexander et al., 2017; Streicher et al., 2018). The FrogCap marker set is modular, such that subsets of the markers can be selected based on the probe-set size, type of research question, and the taxonomic scale. We selected sets of orthologous markers that capture well within and across major lineages of frogs (Hyloidea, Archaeobatrachia, and Ranoidea). From these markers, we designed two complementary probe-sets (referred to as "Hyloidea-V1" and "Ranoidea-V1;" Archaeobatrachians, that represent a smaller subset of frogs are included in Hyloidea-V1; Table 1). The complementary design allows for variable markers to be combined across these two superfamilies (Feng et al., 2017). We also tested the modularity of FrogCap by creating a third probe-set, which is a reduced version of Ranoidea-V1 using half the number of probes ("Reduced-Ranoidea"), to evaluate whether the reduction in probes leads to greater capture success and increased sequencing depth of coverage.

We evaluated these three probe-sets using 121 samples of new sequence data from the broadest phylogenetic scale of Anura down to different populations of a species (Table 2). We evaluated the sensitivity (base-pairs of target markers successfully captured) and specificity (percent reads that map to targets) of the probe design, the number of markers captured, and depth of coverage across the different probe-sets. Because of the expansive phylogenetic breadth of the FrogCap probe design and sampling, we can explore variation in characteristics (alignment length, parsimony informative sites [PIS], and missing data) of different markers at different phylogenetic scales. We examined the effects of phylogenetic relatedness on our capture success, assessing how genetic distance relates to missing data, sensitivity, and specificity of the sequence capture. Lastly, we tested whether using fewer probes in Reduced-Ranoidea leads to greater capture success and increased depth of coverage from the increased sequencing effort on fewer markers.

TABLE 1 Marker contents targeted for the three probe-sets designed in this study

	Hyloidea-V1	Hyloidea-V2	Ranoidea-V1	Ranoidea-V2	Reduced
N probes	40,040	40,040	40,040	40,040	20,020
N base-pairs targeted	2,929,956	3,048,207	3,454,114	3,313,548	1,519,233
N markers targeted	9,229	10,633	13,517	12,909	3,247
Hyloidea-V1 overlap	–	8,194	6,130	6,931	1,652
Hyloidea-V2 overlap	8,194	–	7,823	8,788	2,314
Ranoidea-V1 overlap	6,130	7,823	–	10,476	3,136
Ranoidea-V2 overlap	6,931	8,788	10,476	–	2,953
Reduced overlap	1,652	2,314	3,136	2,953	–
N exons	6,977	8,548	12,834	10,743	3,161
N multiexon genes	986	2,414	2,606	2,541	236
N genes	5,216	5,774	8,132	6,872	3,009
N UCEs	2,166	2,085	651	2,080	0
N Legacy	86	86	32	86	86
Mean marker length	317.2	286.7	255.5	256.7	467.8
Maximum marker length	11,549	11,549	11,429	11,429	11,429
Markers <500 bp	6,667	9,737	12,733	12,222	2,554
Markers 500–1,000 bp	2,562	551	616	441	447
Markers >1,000 bp	390	345	168	246	246

Note: Hyloidea-V2 and Ranoidea-V2 was not explicitly tested; however, they represent a revision where failed markers were discarded in favour of additional UCEs and Legacy markers previously used in Sanger sequencing in frogs. The row "*N* genes" below is the number of genes with one or more exon, where "*N* multi-exon genes" only count genes with two or more exons.

TABLE 2 The phylogenetic scale and sampling used for this study, where 121 individual samples have been sequenced and configured into different data sets

Phylogenetic scale	Clade	*N* samples	*N* markers	Probe-set
Order	Anura	48	6,140	Ranoidea-V1 + Hyloidea-V1
Superfamily	Hyloidea	24	9,229	Hyloidea-V1
Superfamily	Ranoidea	24	13,517	Ranoidea-V1
Family	Mantellidae	8	13,517	Ranoidea-V1
Genus	*Cornufer*	24	13,517	Ranoidea-V1
Genus	*Occidozyga*	30	3,247	Reduced-Ranoidea
Species	*C. vertebralis*	16	13,517	Ranoidea-V1

Note: We note that five samples overlap in some scales (i.e., Genus and Superfamily see Table S2). The Order Anura uses shared markers from the Ranoidea-V1 and Hyloidea-V1 probe-sets from the two Superfamily scales, and includes 24 samples from the Hyloidea data set and 24 samples from the Ranoidea data set.

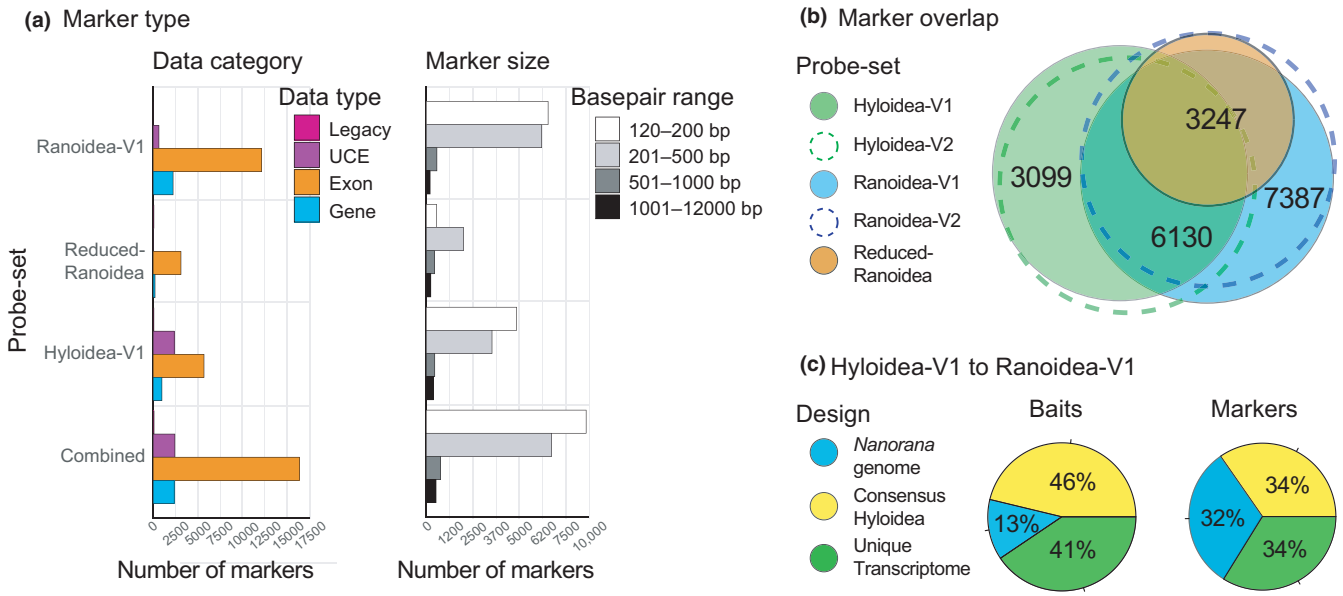

FIGURE 1 The modularity of the FrogCap probe-set permits the selection of different types of markers. The data category in (a) shows the quantity of different marker types (Legacy, UCE, Exon, Gene) used in the design of each probe-set (40,040 baits used for Ranoidea-V1 and Hyloidea-V1; 20,020 baits used for Reduced-Ranoidea). The marker size distribution also shows the general size classes of markers used for each of the probe-sets. The "Combined" probe-set refers to the number of unique markers across all probe-sets to represent the total available markers for FrogCap. In (b) a Venn diagram shows the marker overlap between the probe-sets, where the dotted lines indicate each V2. In (c) the percentage of baits and markers are shown for each type of design: (a) "Nanorana genome" was the *Nanorana parkeri* genome sequence from Ranoidea-V1; (b) "Consensus Hyloidea" which were Hyloidea consensus sequences of the Ranoidea-V1 target marker; and (c) "Unique Transcriptome" which were markers obtained from Hyloidea transcriptomes and are not present in Ranoidea-V1. Different configurations can be created based on bait kit size, modules, marker size or phylogenetic group

2 | MATERIALS AND METHODS

Detailed methods for all sections can be found in the Appendix S1.

2.1 | Sequence capture probe design

2.1.1 | Ranoidea genome-based design

To target exons for the probe-set, we selected orthologous, protein-coding exons that were well represented across the superfamily Ranoidea (AmphibiaWeb, 2021). We conducted all data processing in R (RDCT 2018), using customized scripts with the following R packages: GENOMICRANGES (Lawrence et al., 2013), SEQINR (Charif & Lobry, 2007), and APE (v5.0; Paradis & Schliep, 2019). We used the *Nanorana parkeri* genome and annotations (Sun et al., 2015; available at GigaScience: dx.doi.org/10.5524/100132). We also assembled 18 Ranoidea transcriptomes obtained from previously published studies (Table S1) with BRIDGER (Chang et al., 2015), trimming adaptors with PEAT (Li et al., 2015). We reduced redundancies in the assemblies with CD-HIT (Li & Godzik, 2006). To identify exons with the genome annotations, we used the program BLAT (-minIdentity = 65,

default settings; Kent, 2002) to match transcripts of each assembled transcriptome to the *Nanorana* genome. The sequences of candidate exons were aligned to the *Nanorana* genome sequence using MAFFT v7.312 (-auto; Katoh & Stanley, 2013). To assist in exon selection, alignments were assessed and filtered to include orthologous markers with broad phylogenetic representation. The *Nanorana* sequence for each candidate exon was used to design a MYbaits-2 (40,040 baits) custom bait library (Arbor Biosciences), using 120 mer baits tiled at 2× with 1× tiling 60 bp into the intron. For UCEs, we used the 120 mer UCE Amniote probes from Faircloth et al. (2012) (https://www.ultraconserved.org). Ranoidea-V1 included 13,517 markers covering 3,454,114 bp (Table 1; Figure S1). After testing Ranoidea-V1 in this study, we created a revised Ranoidea-V2 that excludes uncaptured markers and includes 64 additional Legacy Sanger markers from Feng et al. (2017) and longer exons (see below; Table 1).

2.1.2 | Hyloidea transcriptome-based design

In contrast to Ranoidea, a genome was not available for Hyloidea when the study was designed. To design probes from orthologous markers, we used a transcriptome-based approach. We assembled 12 transcriptomes representing six Hyloidea families and one Archaeobatrachia family from previous studies (Table S1). Next, we used the exons designed from Ranoidea-V1 and BLASTed them to the transcriptomes. When the exon matched >five transcriptomes with <20% divergence, the same baits from Ranoidea-V1 were used for Hyloidea-V1 which totaled 5,159 baits (Figure 1c). When there was >20% divergence, new consensus sequences were made from the Hyloidea alignments consuming 18,054 baits. Consensus sequences were generated from transcriptome alignments using the majority base per column or with the base that would maintain optimal GC content if base frequencies were equal. The remaining 15,834 baits were used on transcriptome-based Hyloidea exons. To add additional exons, we used VSEARCH (Rognes et al., 2016) to cluster orthologous transcripts from the Hyloidea transcriptomes and followed the general procedure for marker selection described for Ranoidea-V1. We used the divergent Archeobatrachian frog *Xenopus* genome (Hellsten et al., 2010) to evaluate paralogues, and removed any markers with multiple matches. Finally, we designed probes as described for Ranoidea-V1 using 40,040 baits. After initial synthesis at Arbor BioSciences, 4,055 baits failed synthesis and instead a 90 mer bait was synthesized using the central sequence from the 120mer bait. The final Hyloidea-V1 probe-set targeted 9,229 markers covering 2,929,956 bp (Table 1; Figure S1). Although not tested, we created a Hyloidea-V2 that excluded 1,035 markers found in less than half the samples from this study and added 2,439 new exons. Hyloidea-V2 targets 10,633 markers covering 3,048,207 bp (Table 1).

2.1.3 | Reduced-Ranoidea marker selection

To test the modularity of FrogCap, we selected a subset of markers from Ranoidea-V1 for a smaller 20,020 bait kit. Reduced-Ranoidea was designed after testing Ranoidea-V1; we excluded markers not captured in the target groups (*Occidozyga* and *Kaloula*, families Dicroglossidae and Microhylidae respectively). We added 86 new markers from Feng et al. (2017), described below. We also included 47 ultra-long exons (>5,000 bp) previously excluded from Ranoidea-V1 (Table 1; Figure 1). Reduced-Ranoidea included 3,247 markers targeting 1,519,233 bp of data (Table 1; Figure S1).

2.1.4 | Previously published legacy markers

To maintain compatibility with previous studies, we selected 36 common nuclear markers (i.e., "Legacy") used in Sanger sequencing phylogenetic studies in frogs (Frost et al., 2006; Pyron & Wiens, 2011) for Ranoidea-V1. We also included 86 Legacy markers from Feng et al. (2017) in Hyloidea-V1, Ranoidea-V2 and Reduced-Ranoidea. Finally, to incorporate mitochondrial Sanger markers, we provide a pipeline to extract mitochondrial genomes and separate them into different markers for alignment (see below).

Nuclear marker sequences were designed from the consensus sequences across the multiple sequence alignments from Feng et al. (2017) and used for probe design. For UCEs used in previous frog studies we selected a subset previously sequenced from *Kaloula* (Alexander et al., 2017) captured in >50% of the samples and with >10% PIS, for a total of 651 UCEs in Ranoidea-V1. Improving upon this for Hyloidea-V1 and Ranoidea-V2, we included the 2,166 UCEs from Streicher et al. (2018), which contain the 651 UCEs selected from *Kaloula*. For these UCEs, we redesigned probe sequences by creating consensus sequences across the multiple sequence alignments for each UCE.

2.1.5 | Marker and bait overlap

An important design component of the probe-sets is that they contain overlapping markers (Table 1; Figure 1) which can be reused and combined in future studies. For phylogenetic markers, Hyloidea-V1 and Ranoidea-V1 share 6,130 markers (Figure 1b) and Ranoidea-V2 shares 6,931 markers with Hyloidea-V2. For baits, Hyloidea-V1 used 5,159 Ranoidea-V1 baits with <20% divergence from Hyloidea transcriptomes (Figure 1c). When markers had >20% divergence, new consensus sequences were made from the Hyloidea-V1 sequences totaling 18,054 new baits. The remaining 16,827 baits were used in the Hyloidea transcriptome-based markers.

2.1.6 | Taxon sampling and library preparation

We explored the performance of the probe-sets across seven datasets with different taxon sampling schemes (Table 2). We selected 121 samples configured into the following phylogenetic scales (with samples shared among scales; Table S2): (1) At the level of Anura Order, 48 samples from Ranoidea-V1 and Hyloidea-V1 (24 each);

(2) Superfamily, 24 samples from each superfamily (the same samples used in Order-level; Hyloidea-V1 also tests out one Salamander and two Archaeobatrachrian frogs); (3) Family, eight samples from eight genera in the family Mantellidae (Glaw & Vences, 2006) using Ranoidea-V1; (4) Genus, 24 species sampled broadly from the genus *Cornufer* (Ceratobatrachidae; Brown et al., 2015) using Ranoidea-V1; and (5) Species, 16 samples from one species (*Cornufer vertebralis*) from four Solomon Islands insular populations. Reduced-Ranoidea was evaluated using 30 Philippine *Occidozyga* sampled from throughout the archipelago (Chan et al., 2021).

We extracted DNA from these tissue samples and genomic libraries were prepared by Arbor BioSciences. Finally, samples were sequenced on an Illumina HiSeq 3000 with 150 bp paired-end reads. We acknowledge that our analyses rely upon capture experiments done a single time and parameters of the experiment (e.g., timing, sequencing centre, library preparation location, technicians) could lead to biased and unexpected results.

2.2 | Data processing and alignment

2.2.1 | Data processing pipeline

A bioinformatic pipeline for removing adapter contamination, assembling contigs, and exporting alignments is scripted in R and available at GitHub (bioinformatics-pipeline-v1; https://github.com/chutter/FrogCap-Sequence-Capture). Raw reads were cleaned of adapter contamination, low complexity sequences, and other sequencing artifacts using the program FASTP (Chen et al., 2018). Adapter-cleaned reads were screened against a database of common contaminants (bacteria, human, model organisms, Laurence et al., 2014; see Table S3). Next, we merged overlapping paired-end reads with BBMerge and removed duplicates using "dedupe" using BBTools (Bushnell et al., 2017). Merged singletons and unmerged paired-end reads were assembled de novo using SPADES v.3.12 (Bankevich et al., 2012) applying different k-mer values. DIPSPADES (Safonova et al., 2015) was used to assemble contigs that were polymorphic by randomly selecting one of the two possible bases. Finally, contigs were annotated by matching to the probe-set reference sequences using BLAST (dc-megablast).

The final sets of matching markers were aligned using MAFFT (settings: local-pair; max iterations = 1,000; ep = 0.123; op = 3; --adjust-direction). We screened each alignment for samples ≥40% divergent from consensus sequences, which were almost always incorrectly assigned contigs, chimeric assemblies, or misaligned sequences. Alignments were retained if they included ≥four taxa and were ≥100 bp long. Alignment were externally trimmed until the sites on each end were represented by ≥50% of the samples. Alignments were next separated into different data sets: (1) "Exons," identified exons were adjusted to be in an open-reading frame in multiples of three bases and trimmed to the largest reading frame that accommodated >90% of the sequences; (2) "Introns," consensus sequences of the previous Exon data set were aligned to the original contig and the exon region was removed and the two intron ends were concatenated; (3) "UCEs," were separately saved and not modified; (4) "Legacy," markers (defined above) were saved separately; (5) "Gene," Exon alignments from above were concatenated when they were found to be from the same predicted gene in the *Nanorana* genome to generate longer alignments from linked genes. Finally, Intron and UCE alignments were trimmed using trimAl (*automatic1*; Capella-Gutierrez et al., 2009). We note that Introns could potentially contain UTRs, but without a well-annotated genome it is difficult to distinguish them.

2.2.2 | Mitochondrial genomes

To maintain compatibility of FrogCap with the published mitochondrial Sanger data, we developed a pipeline to extract mitochondrial genomes in R (mitogenome-pipeline-v1; https://github.com/chutter/FrogCap-Sequence-Capture). We tested the pipeline on the Order-level phylogenetic data set of 48 samples (24 from Ranoidea; 24 from Hyloidea). First, we used BLAST to match the *Nanorana* mitochondrial genome (GenBank Accession: NC_026789) to the sample contigs. If there were duplicate matches, we kept the match with the highest bit-score. Finally, we extracted the sequence data for each mitochondrial marker and separated them into a fasta file to be aligned with the bioinformatics-pipeline-v1. With the mitochondrial genome assemblies, markers can be extracted and aligned to barcode and other sequences available on GenBank.

2.3 | Sequence capture evaluation

2.3.1 | Sequence capture sensitivity

We evaluated the "sensitivity" of the capture results, where sensitivity (i.e., "breadth of coverage") is defined as the percent bases from target markers used in probe design covered by post-assembly contigs. To calculate sensitivity, we used target markers from Hyloidea-V1 (n = 22 samples; excluding the two samples with less than 50% capture success), Ranoidea-V1 (24), and Reduced-Ranoidea (30) and BLAST to match them to the sample contigs. We calculated the percent sensitivity per target marker and per sample by dividing the base-pair length of the target marker sequence by the length of the matching portions of the sample contig.

2.3.2 | Sequence capture specificity

"Specificity" refers to the percentage of cleaned reads that can be mapped back to the target markers (number mapped reads/total cleaned reads). We assessed specificity within Hyloidea-V1 (n = 22 samples), Ranoidea-V1 (24), and Reduced-Ranoidea (30). We created an indexed reference from the target markers and mapped cleaned reads from each sample using the program BWA

v0.712 (functions: bwa index and bwa mem; Li et al., 2009). We used SAMTOOLS (Li et al., 2009) to convert between file-types (functions: view and fastq), and counted the number of the reference-mapped cleaned reads to calculate specificity.

2.3.3 | Sequence capture missing data

To assess missing data from the different probe-sets, we characterized variation in two ways: (1) missing base-pair data (ms-bp), the percent of base-pairs missing for a sample (excluding indels) in the trimmed alignments before separating into data types such that all alignments were included in a single data set; and (2) missing marker data (ms-marker), percent of markers missing for a sample in these trimmed alignments.

2.3.4 | Effect of genetic distance

We performed linear regressions to determine if genetic distance is a significant predictor of sensitivity, specificity, and ms-marker. We compared the genome-designed Ranoidea-V1 and consensus-based design Hyloidea-V1. Genetic distance was calculated using uncorrected pairwise distance and the mean was computed across markers for each sample; for Ranoidea-V1, distance was calculated from the *Nanorana* genome sequence, whereas in Hyloidea-V1 distance was calculated from the target marker consensus sequences as described above. For these analyses, we included 18,054 baits from markers shared between the probe-sets that used the consensus-based design in Hyloidea-V1. A significant negative relationship between sensitivity and specificity and genetic distance supports the hypothesis that capture success is driven by sample distance from the target markers. A significant positive relationship between genetic distance and missing data would suggest that sample dissimilarity from the design markers leads to missing marker data.

2.3.5 | Marker depth of coverage

The "depth of coverage" or "depth" was calculated for each sample and marker, where depth refers to the number of bases from the cleaned reads overlapping with an assembled base or bin of bases (often denoted as "X"). We first created a reference for each sample's set of post-assembly contigs targeted with the probe-set and mapped cleaned reads to these contigs using BWA ("bwa-mem" function). Next, we removed exact duplicate reads using PICARD (http://broadinstitute.github.io/picard/). To calculate per-base overlap of cleaned reads to contig base-pairs, the "depth" function was used from SAMTOOLS. Depth was calculated across all targeted markers and samples for every base-pair and was binned into 1% sized bins, calculated for the exons and introns separately. To compare depth values between samples we normalized depth by using reads per kilobase per million (RPKM) to compare among samples with uneven sequencing effort through scaling by gene length and mapped reads (Mortazavi et al., 2008). We also counted overlapping reads in one-percent bins and calculated the RPKM for each bin by scaled by gene.

To describe the variation in depth among samples and markers, we calculated two metrics: (1) sample depth, median depth of markers calculated for each sample; and (2) marker depth, for each marker, median depth of samples calculated for that marker. We used median values, because individual depth measurements are not centred on zero and have a positive skew. We used a Student's two-sample *t* test to compare depth of coverage from Ranoidea-V1 and Reduced-Ranoidea.

2.3.6 | Phylogenetic scale

We evaluated the Exon, Intron, UCE, and gene data sets for criteria considered informative in modern, model-based phylogenetic inference (e.g., Graybeal, 1994; Townsend, 2007) for each phylogenetic scale (Table 2). We calculated statistics for each marker: number of taxa, alignment length (bp), percentage missing base-pair data (percent of missing bases), percentage missing sample data (number of missing taxa), number and percent PIS. We performed linear regressions to determine if alignment length was a significant predictor of PIS, using trimmed data sets for the different phylogenetic scales. These metrics were calculated using the Alignment Assessment tool from Portik et al. (2016).

2.3.7 | Population genetics

To assess whether FrogCap could be used for population genetics, we located high quality variants and SNPs. We note that these statistics depend on the sequencing platform, multiplexing strategy, and number of baits used during hybridization enrichment. Therefore, it would be difficult to compare results across different study designs, and we only seek to demonstrate feasible results if researchers follow a similar design. We used GATK v4.1 (McKenna et al., 2010), following best practices for discovering and calling variants (Van der Auwera et al., 2013). We used the GATK phred-scaled quality scores to filter SNPs, where a score of 20 (99% confidence) was considered "strongly supported." Variant calling for SNPs was conducted through a custom pipeline in R (variant-pipeline-v1; https://github.com/chutter/FrogCap-Sequence-Capture).

3 | RESULTS

3.1 | Modular marker selection

An R script to configure and select different probe-sets using the FrogCap marker database provided on GitHub (Custom-Probe-Design; https://github.com/chutter/FrogCap-Sequence-Capture) and the FrogCap website is provided and contains direct downloads

for the different probe-sets (https://www.frogcap.com). The GitHub page provides the preconfigured probe-sets from this study (Table 1; Figure S1). In addition, the configuration script allows the creation of customized probe-sets, which can be filtered and modified to include different clades of interest (based on if they were captured in each clade), different marker types, and different marker lengths. We note that the initial pool of samples used for filtering is contained to this study, so any clades that are missing markers could be due to stochastic reasons unrelated to divergence from the probe sequence. More species and clades will be added to make the database more robust as they are published.

3.2 | Sequence capture evaluation

We sequenced 121 samples, resulting in a mean 1,234 ± 577 mega base-pairs (Mb; range: 466–4,321 Mb) and a mean 8,172,461 ± 3,827,207 (range: 3,090,636–28,621,450) paired reads for each sample (Table S4; Figure S2). Filtering raw reads to remove exact duplicates, low complexity and poor-quality bases, adaptor and contamination from nontarget organisms resulted in a mean 84.5 ± 11% of reads (range: 27%–96%) passing the quality filtration steps. After assembly, our samples yielded a mean 15,832 ± 5,575.9 (range: 6,968–43,113) contigs, with a mean length of 860 ± 92 (range: 128–24,355) bp (Table S4; Figure S2).

The assembled contigs from each sample were matched using BLAST to the target markers from each of the three probe-sets. Hyloidea-V1 had a mean of 7,443 ± 2,022 (range: 616–9,570) contigs that matched uniquely to target markers (mean marker proportion: 0.701 ± 0.199; range: 0.058–0.901; Table S5). Ranoidea-V1 produced a mean of 10,304 ± 1,645 (range: 5,050–12,235) contigs that matched uniquely to target markers (mean marker proportion: 0.757 ± 0.121; range: 0.371–0.898). Finally, Reduced-Ranoidea produced a mean 2,847 ± 123 (range: 2,299–2,983) contig matches (mean marker proportion: 0.877 ± 0.038; range: 0.708–0.919; Table S5).

3.3 | Sequence capture sensitivity

Sequence capture "sensitivity" was measured across our three probe-sets (Hyloidea-V1, Ranoidea-V1, Reduced-Ranoidea) by assessing the percent base-pairs of target markers covered by post-assembly contigs. The mean sensitivity across all markers in Hyloidea-V1 was 71.5 ± 20.2% (n = 24; range: 5.8%–91.1%); Ranoidea-V1 had a mean sensitivity of 79.6 ± 15.3% (n = 24; range: 49.8%–94.4%); and Reduced-Ranoidea had a mean sensitivity of 89.4 ± 4.1% (n = 30; range: 71.3%–93.8%; Table 3; Figure 2a). We used only shared markers to compare Hyloidea-V1 and Ranoidea-V1, where Hyloidea-V1 had a mean sensitivity of 73.7 ± 12.9% (n = 22; range: 40.8%–91.0%) and Ranoidea-V1 had a mean sensitivity of 81.5 ± 14.1% (n = 24; range: 52.5%–95.1%).

3.4 | Sequence capture specificity

Specificity, defined as the percentage of cleaned reads mapped to the target markers, was assessed within Hyloidea-V1, Ranoidea-V1, and Reduced-Ranoidea. Hyloidea-V1 had a mean specificity of 37.8 ± 9.2% (n = 24; range: 1.7%–54.3%); Ranoidea-V1 had a higher mean specificity of 21.4 ± 7.5% (n = 24; range: 7.4%–32.1%); and specificity was lowest in Reduced-Ranoidea, with a mean of 20.2 ± 3.3% (n = 30; range: 12.2%–23.6%; Table 3; Figure 2b). We used only shared markers to compare Hyloidea-V1 and Ranoidea-V1, where Hyloidea-V1 had a mean sensitivity of 18.7 ± 6.6% (n = 22; range: 6.2%–32.4%) and Ranoidea-V1 had a mean sensitivity of 12.5 ± 4.5% (n = 24; range: 4.2%–19.1%).

3.5 | Sequence capture missing data

We assessed missing data for the three probe-sets (Hyloidea-V1, Ranoidea-V1, Reduced-Ranoidea) by calculating the percent of missing base-pairs (ms-bp) and missing markers (ms-markers) from alignment lengths and target marker counts for each sample, respectively (Figure 3a,b). The mean ms-bp from samples in Hyloidea-V1 was 29.7 ± 20% (n = 24; range: 10.1%–94.3%), with a larger range and variation because of the inclusion of divergent clades (Archaeobatrachian frogs and Salamanders). Ranoidea-V1 mean ms-bp was 29.4 ± 19% (n = 24; range: 6.9%–63.1%). Last, Reduced-Ranoidea mean ms-bp was 10.9 ± 4% (n = 30; range: 6.5%–28.4%; Table 3; Figure 3a). For ms-markers, Hyloidea-V1 had a mean 21.9 ± 22.7% (n = 24; range: 0%–95.0%). In Ranoidea-V1, mean ms-markers was slightly higher at 22.7 ± 21.2% (n = 24; range: 0%–59.7%). Finally, ms-markers in Reduced-Ranoidea were lower with a mean 4.5 ± 4.2% (n = 30; range: 0%–23.2%; Table 3; Figure 3b).

3.6 | Effect of genetic distance

When testing for sensitivity in genome-designed Ranoidea-V1, we found a significant negative relationship between genetic distance and sensitivity (n = 24; R^2 = .689; p < .001; Figure 2a), as described by the equation [sensitivity = 1.065 + −0.249 * mean pairwise divergence], with pairwise divergence and sensitivity measured as percentages. Sensitivity decreased 0.249% for each percent increase of pairwise divergence. Conversely, we found a nonsignificant relationship among Hyloidea-V1 consensus sequence samples (n = 22; R^2 = .133; p = .095; Figure 2a), which was not impacted by log-transforming input values.

We found a significant negative relationship between genetic distance and specificity (R^2 = .837; p < .001) among genome-designed Ranoidea-V1 samples (Figure 2b), described by the equation (specificity = 0.630 + −1.59 * mean pairwise divergence). Specificity decreased 1.59% for each percent increase of pairwise divergence. We did not find a significant relationship between genetic distance

TABLE 3 Sequence capture evaluation for the three probe-sets

	Hyloidea-V1	Ranoidea-V1	Reduced-Ranoidea
N probes/baits	40K	40K	20K
N base-pairs	2,929,956	3,454,114	1,519,233
N markers	9,229	13,517	3,247
Mean sensitivity (%)	70.3%	70.6%	89.1%
Mean specificity (%)	47.9%	71.1%	35.7%
Mean marker missing data (%)	21.9%	22.7%	4.5%
Median depth of coverage (RPKM)	52.3	47.3	158.1
Median exon depth (RPKM)	47.3	52.3	158.1
Median intron depth (RPKM)	21.1	23.2	80.8
Median UCE depth (RPKM)	47.2	50.6	-
Median gene depth (RPKM)	52.9	47.0	156.6

Note: Reduced-Ranoidea uses the same markers from Ranoidea-V1 (with some additions; see Materials and Methods) but was designed with half the number of baits for a specific taxonomic group (*Occidozyga* and *Kaloula*) is defined as the percent base-pairs of target markers successfully captured and specificity is the percent reads that map to targets.

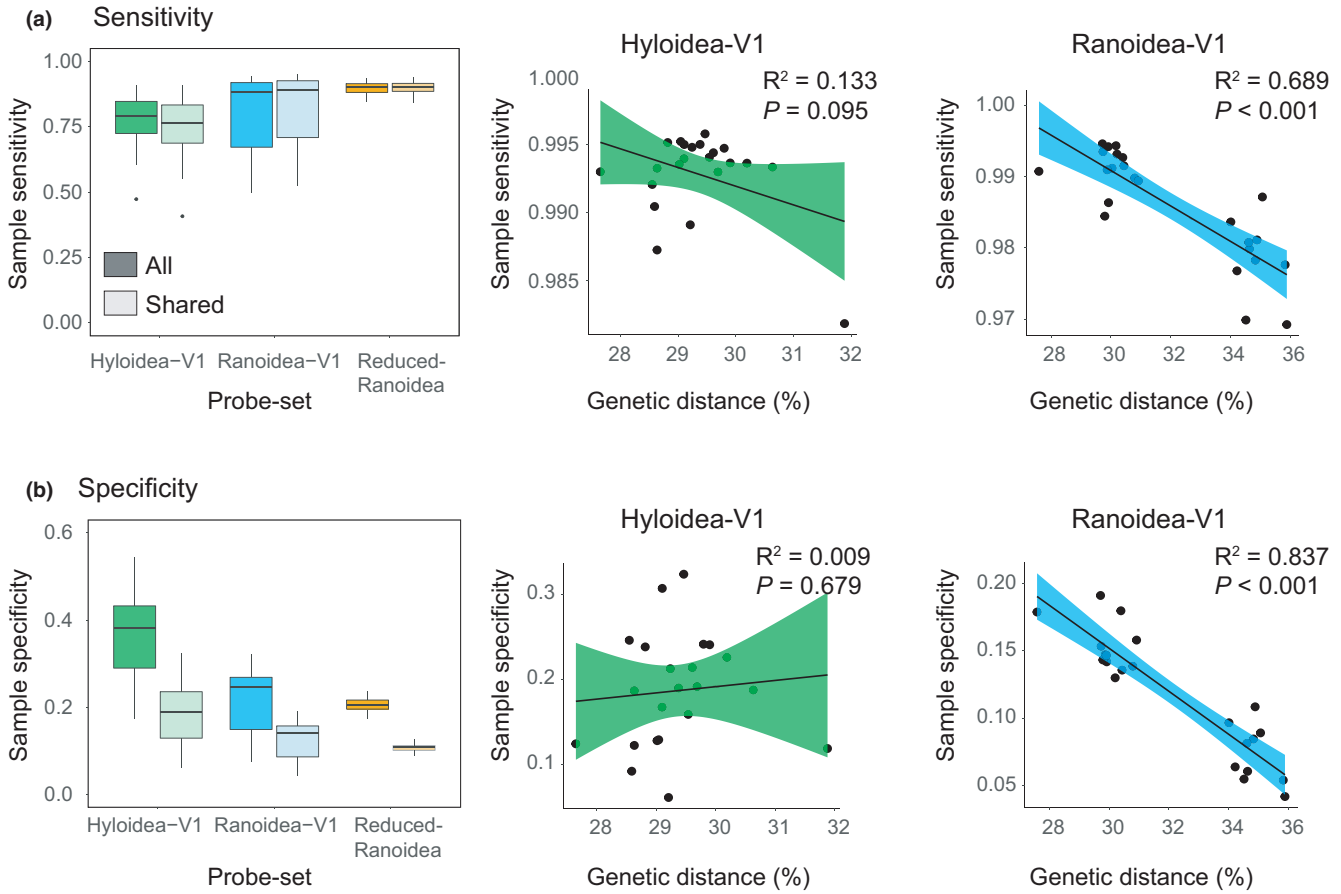

FIGURE 2 Sample sensitivity (percent bases from target markers that are covered by post-assembly contigs) and specificity (percent cleaned reads mapped to the target markers) are compared between shared markers in the Hyloidea-V1 and Ranoidea-V1 probe-sets (Reduced-Ranoidea was designed differently and is not comparable). The darker coloured box plots use all their target markers while the lighter coloured box plots show markers overlapping between Ranoidea-V1 and Hyloidea-V1. In addition, we evaluate relationships between genetic distance and sensitivity/specificity for Hyloidea-V1 and Ranoidea-V1 (Reduced-Ranoidea not included because samples are all from a single genus and have similar genetic distances). The box plots show the distribution of (a) sensitivity and (b) specificity values across the probe-sets. Sensitivity has a significant negative relationship with genetic distance in Ranoidea-V1 and is nonsignificant in Hyloidea-V1. Specificity has a significant negative relationship in Ranoidea-V1 and a nonsignificant relationship in Hyloidea-V1. The 95% confidence intervals of the estimated regression line are indicated by coloured shading

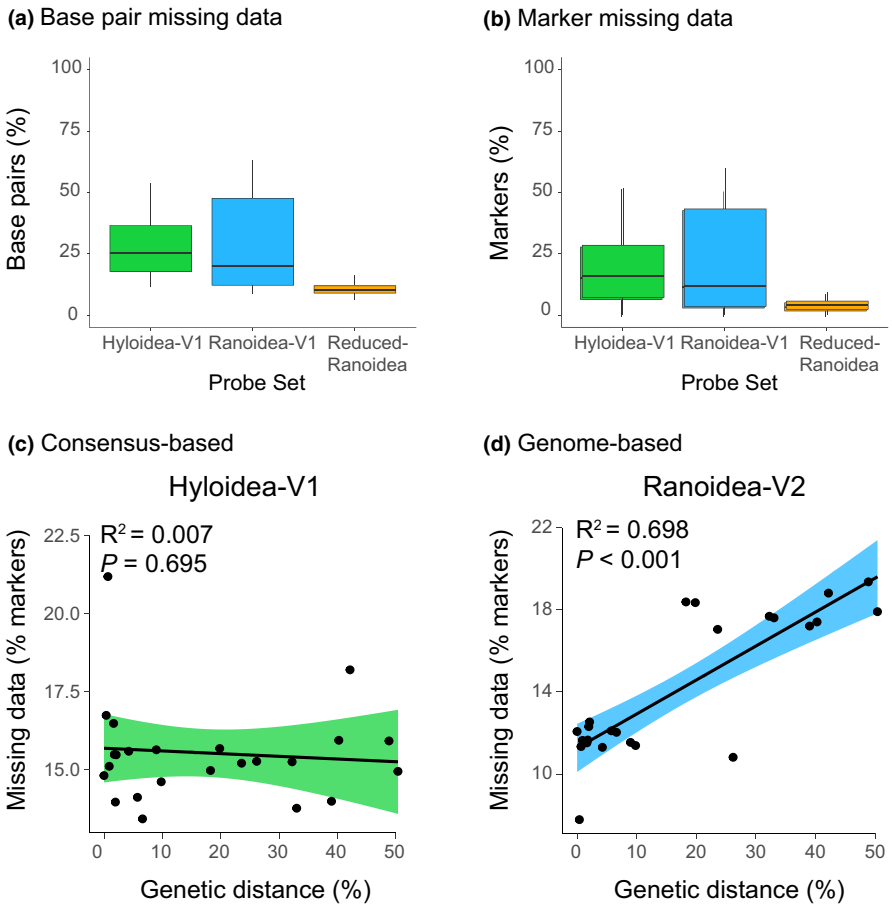

FIGURE 3 Missing data are compared for the three probe-sets (Hyloidea-V1, Ranoidea-V1, Reduced-Ranoidea). Base-pair missing data in (a) is the percent of missing data calculated from the number of missing bases pairs across alignments for each sample. Marker missing data in (b) is the percent of missing markers across alignments for each sample. The consensus-based marker design missing data from Hyloidea-V1 in (c) has a nonsignificant relationship with genetic distance. The genome-based missing data from Ranoidea-V1 in (d) has a strong and significant positive relationship with genetic distance. Reduced-Ranoidea is not included because samples are all from a single genus and have the same genetic distances. The 95% confidence intervals of the estimated regression line are indicated with coloured shading

and specificity within Hyloidea-V1 consensus sequence samples (R^2 = .009; p = .679; Figure 2b).

Finally, we found a significant positive relationship between ms-markers and genetic distance (R^2 = .698; p < .001) in genome-designed Ranoidea-V1 samples (Figure 3c). In Ranoidea-V1, ms-markers are equal to the equation (ms-markers = 11.27 + 1.66 * mean pairwise distance), when pairwise distance and missing data are measured as percentages. The ms-markers increased 1.66% for each percent increase of pairwise divergence. In the consensus-designed Hyloidea-V1 samples, we found a weak nonsignificant relationship (R^2 = .007; p = .695; Figure 3d).

3.7 | Marker depth of coverage

We assessed depth of coverage across markers using the post-assembly contigs and then separately for exons targeted with probes, and introns that were incidentally sequenced. We measured depth by counting the number of reads that overlapped each base (denoted as "X"), which we note is not a comparable measure between samples, so RPKM is used for comparison (raw "X" measures shown in Figure S3). In Hyloidea-V1, we found a median depth calculated across samples (i.e., sample depth) of 19.7 ± 6.9× (n = 24; range: 5.1–34.1×; Figure 4a). In Ranoidea-V1 we found a median sample depth of 22.1 ± 8.5× (n = 24; range: 4.2–39.2×; Figure 4a). Reduced-Ranoidea had a median sample depth of 26.8 ± 5.5× (n = 30; range: 14.2–34.7×; Figure 4a). To compare the probe-sets with varying sample sequencing depths, we scaled depth by RPKM and found a median sample depth of 59.7 ± 21.4 RPKM (n = 24; range: 45.9–118.6 RPKM; Figure 4a) in Hyloidea-V1, a median sample depth of 61.3 ± 15.2 RPKM (n = 24; range: 47.9–112.3 RPKM; Figure 4a) in Ranoidea-V1, and a median sample depth of 173.6 ± 16.2 RPKM (n = 24; range: 126.8–212.1 RPKM; Figure 4a) in Reduced-Ranoidea. With a Student's two-sample t test we found a significant difference in the means of the median sample depth between Reduced-Ranoidea and Ranoidea-V1 (t = −19.847, df = 46.4, p < .001). The results for the median depth across markers (rather than samples) and comparisons of depth between introns/exons were similar and can be found in the Appendix S1.

3.8 | Phylogenetic scale

The resulting number of alignments before delimiting into data types (Exons, Introns, UCEs, Legacy, Genes) for the phylogenetic scales was variable (Tables 2 and 4; Figure S4a,b). Reduced-Ranoidea had 3,156 alignments (out of 3,247 targeted markers). The phylogenetic scale with the next fewest alignments was Order-level with 5,500 alignments. The superfamily Hyloidea resulted in 8,712 alignments (because this probe-set intentionally was designed with longer markers), and the remainder

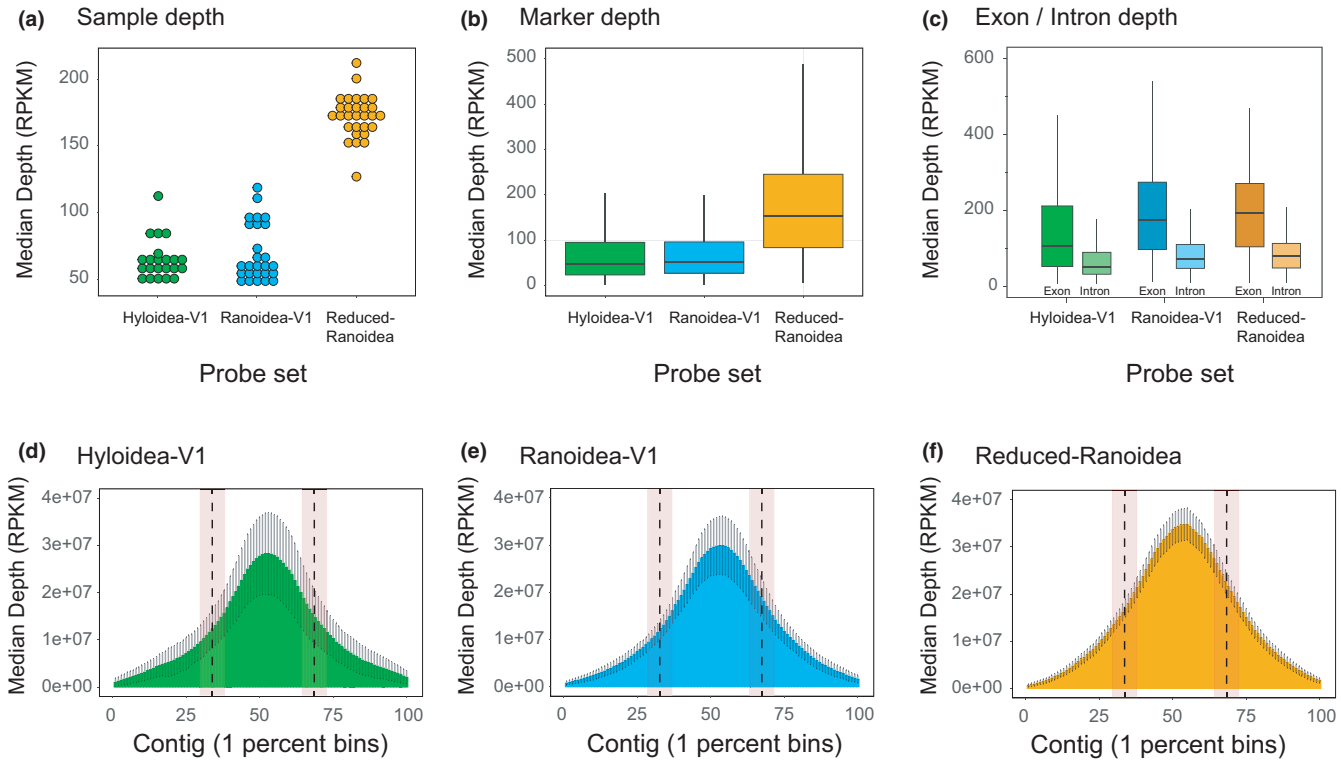

FIGURE 4 Depth of coverage statistics calculated for the three probe-sets. Depth measures use the mapped Reads Per Kilobase Per Million (RPKM) to compare among samples with uneven sequencing effort through scaling by gene size and mapped reads. Depth is summarized for (a) median depth of markers for each sample (the individual points); (b) median depth of markers calculated across all samples for that marker; (c) median depth of markers was calculated for the exon and intron separately; and in (d–f) depth was calculated for each sample marker by counting overlapping cleaned reads in one-percent bins and calculating the RPKM for that bin scaled by gene. Each bar is the median depth (RPKM) across all binned sample markers. The grey coloration represents the standard deviation within each 1 percent bin across markers. The vertical dotted lines in each plot give the mean position where exon-intron boundaries occur, with the standard deviation shown in the red-coloured shading

of the scales had similar numbers of alignments at 12,009–13,099 because they resulted from Ranoidea-V1 (Table 4; Figure S4).

Mean alignment length was variable among the phylogenetic scales. Alignments at the Order-level were among the shortest. This can be largely attributed to trimming highly variable intron regions, which also resulted in many short alignments being discarded during filtering steps (Table 4; Figure S4c). Conversely, alignment lengths increased at shallower phylogenetic scales since intron regions were less variable and easier to align, and as a result were not trimmed as aggressively (Table 4). Missing marker data at the broadest phylogenetic scales (e.g., Order and Superfamily) were higher than other scales (Figure S4d).

There was also a general pattern of markers exhibiting increased PIS at broader phylogenetic scales (Figures S3b and S4). Introns had the most PIS across all phylogenetic scales and surprisingly, even displayed variation at the Species-level (Table S5; Figure S5). Exons and Genes (composed of multiple exons) were moderately variable. The UCEs were slightly less variable than the Exons and Genes, but variation occurred in their characteristic flanking regions. Finally, we found a significant relationship between alignment length and percent PIS across all phylogenetic scales. The strength of this relationship decreased at shallower phylogenetic scales (Genus- and Species-level), indicating marker variation generally decreased at shallower scales (Table S5). However, clear differences in variation exist between Exons and Introns, with Introns providing a majority of PIS detected at shallower scales (Table S6; Figure S5).

3.9 | Mitochondrial genomes

To maintain compatibility with the vast mitochondrial Sanger sequencing data available, we provide a pipeline to extract mitochondrial genomes from assembled contigs. We found success in most of the 48 samples tested from the order-level data set, where one sample did not have contigs that matched to the mitochondrial reference while 47 samples matched. The matching samples had a high percent completeness of 83.7 ± 15.9% (range: 49.4%–100%) for the target markers (i.e., not the tRNAs). Across target markers 81.1 ± 13.5% (range: 52.9%–94.6%) of the samples had data for that marker (Figure 5a).

3.9.1 | Population genetics

To assess whether FrogCap has utility in population genetic studies, we used GATK4 to discover genetic variants and SNPs addressing

TABLE 4 Summary of each data set for each phylogenetic scale (Anura, order; Hyloidea and Ranoidea, superfamily; Mantellidae, family; *Cornufer*, genus; *C. vertebralis*, species; and *Occidozyga*, Reduced-Ranoidea)

Scale (probe-set)	Data set	*N* samples (mean ± sd)	*N* markers	Total bp	Length bp	Total PIS	PIS/target (mean ± sd)
Anura (shared)	Exon	34.3 ± 7.9	4,566	961,305	210.5 ± 145.4	409,774	40.4 ± 8.9
	Intron	31.6 ± 7.9	4,185	1,317,305	314.8 ± 171.4	1,041,290	80.5 ± 14.2
	Legacy	42.8 ± 3.5	15	15,258	1017.2 ± 412.2	5,523	35.6 ± 6.7
	Gene	41.6 ± 4.7	857	457,743	534.1 ± 323.7	190,521	40.6 ± 6.4
	UCE	42.3 ± 5.7	567	319,259	563.1 ± 178.8	128,088	39.3 ± 12.3
Hyloidea (Hyloidea-V1)	Exon	16.2 ± 5.2	5,696	1,332,048	233.9 ± 241.3	307,447	20.9 ± 8.6
	Intron	15.6 ± 4.9	5,636	2,310,925	410 ± 199.1	1,174,156	52.5 ± 20.3
	Legacy	21.1 ± 4.1	84	79,113	941.8 ± 321.6	19,900	25.2 ± 6.7
	Gene	19.9 ± 3.1	985	546,951	555.3 ± 376.6	121,821	21.1 ± 6.1
	UCE	20.9 ± 3.2	2,359	1,941,867	823.2 ± 232.2	519,351	26.5 ± 10.3
Ranoidea (Ranoidea-V1)	Exon	17 ± 5	12,149	2,453,364	201.9 ± 149.7	718,950	27 ± 9.2
	Intron	16.1 ± 4.4	2,982	1,104,972	370.5 ± 134.7	673,284	62.4 ± 16.6
	Legacy	20.6 ± 4.4	24	21,186	882.8 ± 415.1	5,162	23.2 ± 7.5
	Gene	21.3 ± 3	2,235	1,347,147	602.8 ± 417	388,411	27.6 ± 6.5
	UCE	20.6 ± 4.7	650	442,500	680.8 ± 173.8	138,803	31 ± 10
Mantellidae (Ranoidea-V1)	Exon	7 ± 1.1	11,339	2,413,164	212.8 ± 190	135,430	5.2 ± 2.8
	Intron	7 ± 1.1	11,314	6,700,598	592.2 ± 206.8	1,050,533	16 ± 7.7
	Legacy	7.5 ± 1.1	29	35,352	1219 ± 787	1,621	4.4 ± 2.2
	Gene	7.8 ± 0.5	2,077	1,275,423	614.1 ± 445.5	68,487	5.2 ± 2
	UCE	7.4 ± 1	604	545,819	903.7 ± 215.8	32,938	5.9 ± 3.1
Cornufer (Ranoidea-V1)	Exon	20 ± 4.8	11,990	2,609,055	217.6 ± 231.3	177,481	6.2 ± 3.4
	Intron	19.9 ± 4.7	11,967	6,046,801	505.3 ± 159.8	1,108,147	18.4 ± 7.8
	Legacy	22.1 ± 4	29	36,594	1261.9 ± 828.8	2,198	5.8 ± 3.7
	Gene	23.2 ± 1.8	2,203	1,368,702	621.3 ± 459.2	85,886	6.1 ± 2.3
	UCE	20.6 ± 5	628	473,984	754.8 ± 167	31,250	6.4 ± 4.1
C. vertebralis (Ranoidea-V1)	Exon	14.1 ± 2.9	11,393	2,507,571	220.1 ± 207	18,347	0.7 ± 1.2
	Intron	14.1 ± 2.9	11,375	5,790,769	509.1 ± 134.8	189,974	3.3 ± 4.1
	Legacy	14.6 ± 2.7	29	36,036	1242.6 ± 826.9	201	0.5 ± 0.4
	Gene	15.7 ± 1	2,100	1,314,084	625.8 ± 460.3	8,657	0.7 ± 0.7
	UCE	14.5 ± 2.7	602	436,702	725.4 ± 127.5	4,564	1 ± 1.8
Occidozyga (Reduced-Ranoidea)	Exon	27 ± 5.3	3114	1,292,427	415.6 ± 678.3	75,773	5.3 ± 2.8
	Intron	26.8 ± 5.3	3,100	1,472,238	474.9 ± 138.5	235,791	16 ± 7.1
	Legacy	28 ± 5.9	85	88,188	1037.5 ± 551.9	3,842	4.2 ± 2.3
	Gene	29.6 ± 1.9	213	156,186	733.3 ± 718.1	8,216	5.1 ± 2.3
	UCE	–	–	–	–	–	–

Abbreviations: bp, base-pairs; *N* samples, number of samples in alignments; PIS, parsimony informative sites.

four phylogenetic scales from six datasets. (Figure 5b). Generally, we found a pattern of decreasing variants (i.e., SNPs, indels) from higher to lower phylogenetic scales (Order to Species-levels), with greater than 10,000 SNPs within the same species after highquality filtering for SNPs (Figure 5c). Finally, in studies for which the independence of SNPs is required, we find at least one strongly supported SNP (and often numerous others) per individual marker, thus permitting thousands of unlinked SNPs in FrogCap data sets, even at the Species-level (Figure S6).

4 | DISCUSSION

We introduce FrogCap, a publicly available database of molecular markers designed for frogs (Anura) that includes several preconfigured nested sequence capture probe-sets (Figure 1). FrogCap contains multiple sequence capture probe-sets designed for capture efficiency in different lineages of frogs, yet which overlap considerably in terms of shared markers. FrogCap is a powerful resource to aid in the collection of genomic data for phylogenomics, systematics,

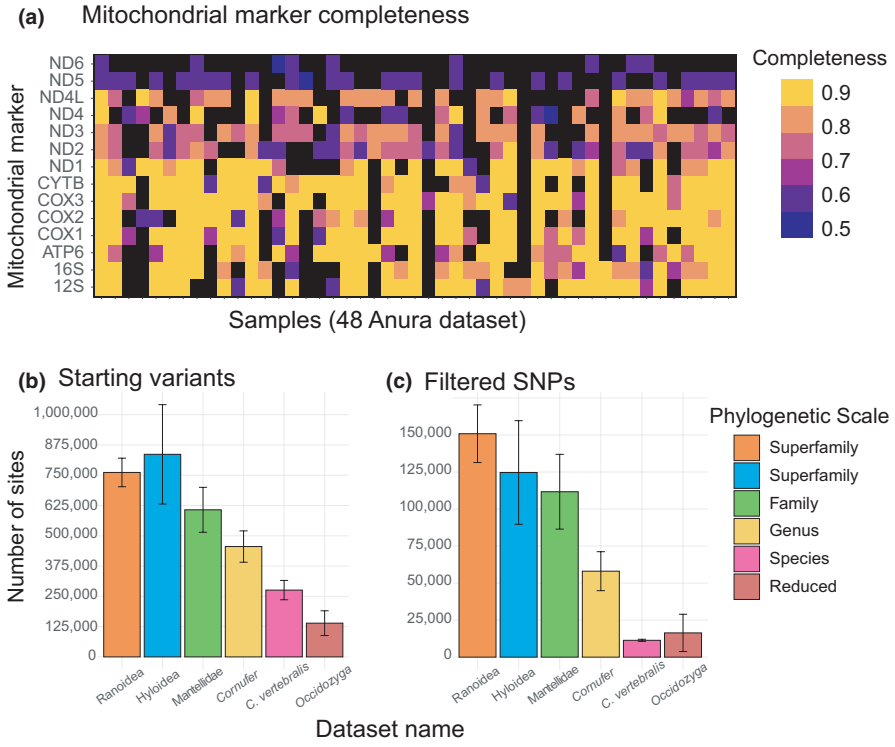

FIGURE 5 Successful extraction results are shown for the non-targeted mitochondrial genomes and number of high-quality SNPs. In (a) the mitochondrial genome marker completeness is shown, with higher numbers (and more yellow colours) indicate greater success in extracting that marker. We excluded tRNAs as they were highly variable and difficult to align and identify. For obtaining SNP data, (b) shows all variants after moderate filtering (quality >5); (c) the number of SNPs after high-quality filtering (phred-scaled quality score >20), if a single independent SNP is required from each marker. In (b) and (c) the mean is shown with the error bars representing two standard deviations from the mean

and population genetics for frog clades ranging in scale from the broadest Anura-level to the smallest population-level questions.

To evaluate the utility of FrogCap, we tested the three main probe-sets (i.e., Ranoidea-V1, Hyloidea-V1, and Reduced-Ranoidea; Table 1) using a sampling scheme of 121 samples that covers five different phylogenetic scales (Table 2). We demonstrate the success of FrogCap at each of these levels. Based on the development and testing of FrogCap, we make several recommendations on key topics related to sequence capture methods. We discuss the effects of using different bait-kit sizes, genome-based versus consensus-based probe design, the impact of phylogenetic scale on sequence capture success, and the importance of incorporating different molecular marker types. Altogether these results suggest that FrogCap is adaptable and provides a much-needed novel mixture of molecular marker types for empirical comparison and downstream filtration which have the potential to resolve difficult phylogenetic relationships.

4.1 | Bait-kit size

One of the first decisions in sequence capture study design is selecting the size of the bait-kit (i.e., the number of probes to purchase). Given the same library pooling design, sequencing effort, and probe tiling scheme, the number of probes available for capture design will directly affect the number of target markers and can indirectly affect the depth of coverage. Here we aimed for an optimal trade-off between depth of coverage and sequencing costs, while maintaining the same conditions for multiplexing sample numbers and sequencing effort. Although larger sized bait-kits can target more markers, they could be disadvantageous if the sequencing effort is not adequate, resulting in higher amounts of missing data. Additionally, a smaller bait-kit that targets a subset of markers would be expected to have greater resulting depth per sample because fewer genomic areas would be targeted for sequencing.

We tested these predictions by using 20,020 versus 40,040 baits in Ranoidea-V1 and Reduced-Ranoidea. Overall, we supported our predictions and found that Reduced-Ranoidea had triple the depth of coverage between shared markers with Ranoidea-V1 (Figure 4). Despite this substantial increase, we found that the 40,040 Ranoidea-V1 samples still had sufficient depth to extract high-quality variants (Figure 5c). With these results, a potential cost-saving measure would multiplex possibly up to triple the number of samples (which was not tested here), which could offer an advantage for projects focused on fewer markers and a reduction in overall sequencing costs.

4.2 | Probe design

Our experiment included probes designed from a combination of transcriptomes and genomes. We found that the genome-based design (Ranoidea-V1), where exon/intron boundaries were clearer, had slightly better capture success than the transcriptome-based (Hyloidea-V1) design. Many of the uncaptured target markers and paralogues detected post-processing in Hyloidea-V1 were transcriptome-based and not included in Ranoidea-V1. This suggests that not properly identifying exon/intron boundaries can lead to poorer capture success. Nevertheless, the transcriptome-based approach successfully recovered ~70% of the target markers, which

is consistent with other transcriptome-based studies (e.g., Bi et al., 2012; Portik et al., 2016; Reilly et al., 2019). Furthermore, the number of genomic resources available for designing Ranoidea-V1 was much higher (Ranoidea: 18 transcriptomes and one genome; Hyloidea-V1: 12 transcriptomes), and probably affected the sequence capture outcomes. More genomes from divergent clades would be necessary to achieve optimal capture success for future probe-sets in frogs.

Probe tiling density is another important variable that differentiates FrogCap from other frog probe-sets. A higher tiling density limits the number of target markers but increases the likelihood of obtaining a target region (e.g., AHE markers; Lemmon et al., 2012). A lower tiling density allows for more target markers but may result in higher levels of missing data. In our study, we used a lower density 2x tiling scheme (at least 50% of each probe overlaps with adjacent probes) that has been recommended as a standard for sequence capture designs (Tewhey et al., 2009). We found a slightly lower capture success (i.e., specificity and missing data) than other studies with more dense tiling designs (specificity: this study = 70%–75%; Portik et al., 2016 = 80%), which led to a higher amount of missing marker data (this study = 20%–25%; Portik et al., 2016 = 8%–10%). However, Reduced-Ranoidea had greater success than these densely tiled studies, indicating that careful selection of markers based on prior successful sequencing can mitigate limitations imposed by lower density tiling. Importantly, our results suggest that the advantages of densely tiled sequence capture designs (fewer markers and more probes) are offset by the inclusion of less dense tiling (more markers and fewer probes) which results in more markers being recovered (Figure 2). Therefore, we recommend less dense tiling schemes (2×) because the number of markers lost from 2× tiling is small compared to the number of markers gained from the larger number being targeted by using fewer probes.

4.3 | Genetic divergence from probes

Genetic distance from the design markers is expected to be one of the most important factors affecting sequence capture performance (including sensitivity, specificity, and missing marker data). Like other studies (Portik et al., 2016), we found higher genetic distances from target markers led to decreased capture success for Ranoidea-V1, which used genome-based marker design rather than consensus sequences (Figure 2). The consensus-sequence probe design for Hyloidea-V1 had poor or nonsignificant relationships between genetic distance and sequence capture performance metrics (Figures 2 and 3). Thus, consensus-sequence probe design appears to remove the negative effects of genetic distance on capture success. These results are comparable to the "common ancestor design" of Hugall et al. (2016) or centroidal sequences from Beaudry et al. (2021). Additionally, Reduced-Ranoidea had overall higher capture success, but is not comparable to the other probe-sets because the markers were selected for their capture success in the groups of interest. Overall, our results suggest that markers designed using consensus sequences improved the sequence capture success when compared to genome-based designs.

4.4 | Phylogenetic scale

Understanding how sequence capture experiments perform at different phylogenetic scales is important for decisions about marker selection and probe design. In our experiment, we assessed the impact of phylogenetic scale on the number of markers recovered, total base-pairs recovered, the number of PIS, and missing data at the levels of Order, Superfamily, Family, Genus and Species (Taylor & Piel, 2004; Townsend, 2007). At broader phylogenetic scales (e.g., Order, Superfamily) the number of alignments decreases because missing data were more prevalent (~25%) because of marker drop-off in genetically divergent samples (Figure S4d). At shallower scales (e.g., Family, Genus, Species), missing data was lower (<10%) and more uniform across samples (Figure S4d). These patterns were especially pronounced for Introns. The high variation of introns at broader phylogenetic scales made aligning homologous regions difficult and resulted in more discarded alignments. At shallower scales, the number of markers recovered appears to be most strongly related to divergence from the design probes. To help remedy this issue, we ensured FrogCap probe-sets are modular. For example, Reduced-Ranoidea was designed based on successful capture data from Ranoidea-V1, where we find a higher capture success and many fewer missing data across alignments.

4.5 | Molecular marker type

We evaluated the properties of UCE, Exon, Intron and gene sequence data within and across phylogenetic scales in this study (phylogenetic performance for these marker types are compared in Chan et al., 2020a). However, intronic sequence was indirectly acquired as "by-catch" from sequencing adjacent regions of a captured DNA fragment (e.g., Bi et al., 2012; Guo et al., 2012; Tewhey et al., 2009). Similar to Portik et al. (2016), the number of base-pairs of data available from nontargeted intronic sequence was 2–3 times higher than explicitly targeted exon sequence (Figure S4), suggesting that intronic sequence is an abundant and potentially important resource in exon-capture. Intronic sequences are regarded as valuable because of more neutral evolution relative to exons which are typically functional and under selection or the core region of UCEs which are probably under strong purifying selection with the flanking regions being more neutral (Halligan et al., 2004; Katzman et al., 2007; Stephen et al., 2008). In assessing the variability (PIS) of different marker classes (Figure S5), we found that Introns are the most variable whereas UCEs and Exons are similar with small to moderate variability. At shallow phylogenetic scales (Species-, Genus-level) Exons and UCEs have little variation, but Introns remain highly variable at all scales (Figure S5). At broader phylogenetic scales (Order-, Superfamily-level), Exons and UCEs become more variable, and variation in Introns become exceedingly high (and leads to alignment removal during trimming and filtering).

The alignment length for exons should be considered when including markers for analyses, and the FrogCap probe-set includes a

mixture of exons from different size classes (Figure S1b). Exon length is a strong predictor of the number of informative sites, and longer exons provide more information for phylogenetic analyses (Table S5; Townsend, 2007). The mean size of the Exons data set was ~250 bp (Table 1), which is near the predicted mean exon length across the *Nanorana* and *Xenopus* genomes (~200 bp). Short exons can be advantageous for population genetics because they allow more SNPs to be incorporated into analyses that require genetically unlinked SNPs (Figure S6); however, they may lack sufficient variability for strong support in phylogenetic analyses, which is important for summary species tree methods relying on individual gene tree topologies. For these types of analyses, FrogCap also includes ~250 large-exons (in Ranoidea-V2 and Hyloidea-V1, V2) greater than 1,000 bp (Table 1), which is comparable to or exceeds the RELEC long-exon set designed for squamates and other amniotes (Karin et al., 2020). Long exons are ideal for analyses where the goal is to have fewer markers/partitions, but stronger statistical support among gene trees (Blom et al., 2017; Maddison, 1997; Richards et al., 2018; Xi et al., 2015). Additionally, we note that selection on protein-coding exons of all sizes may lead to positively misleading topologies (see Introduction). The publicly available FrogCap data processing pipeline separates alignments into these different data types, greatly simplifying the process of filtering data sets for specific analyses.

Finally, an important marker type for species delimitation and population genetics that can be used from FrogCap are SNPs collected from the targeted markers. We show here that FrogCap markers are variable at the Species-level and have been used in a species delimitation study aiding in the description of new species to science (Rasolonjatovo et al., 2020). Using the SNP FrogCap processing pipeline, we found a pattern of decreasing SNPs from higher to lower phylogenetic scales (Order- to Species-levels), with greater than 20,000 variants at the Species-level after filtering to high quality variants (Figure 5). Here, we suggest that there are enough SNPs for population genetic analyses, which is demonstrated in two other demographic studies showing significant utility of FrogCap for demographic analyses (Chan et al., 2020; Chan et al., 2021).

4.6 | Future directions

FrogCap is a collection of molecular markers and sequence capture probe-sets for frogs. The combination of different marker classes targeted by FrogCap for a given study system allows for new research directions and provides a foundation for future work managing phylogenomic data types and analysing whole genomes. Apart from phylogenomics, FrogCap is suitable for other applications such as comparative genomics to assess natural selection across exons. Additionally, as researchers continue to publish FrogCap data sets (Chan et al., 2020, 2020a, 2020c, 2021; Rasolonjatovo et al., 2020), the database of markers can be refined to increase broader capture success as well as aid in probe design for subsets of frog taxa. For example, Ranoidea-V2 and Hyloidea-V2 probe-sets have been refined from these results by replacing markers that captured poorly across the sampling. Finally, sequence capture is not the most direct method for acquiring SNP data (vs. RADseq), yet we demonstrate that thousands of high-quality SNPs can be obtained with FrogCap (Chan et al., 2020). This is important for future studies, as the main disadvantage of RADseq is that the data cannot be easily combined in future studies; sequence capture can provide similar quality data, in addition to using known markers, which ensures that these data can be integrated into future, larger scale phylogenomic studies.

ACKNOWLEDGEMENTS

We would like to thank Travis Glenn and three anonymous reviewers for comments that greatly improved earlier drafts of this manuscript. For funding and financial support that made this research possible, we would like to thank KU Research Investment Council (RIC) Level II Award (No. 2300207) to RMB and R. G. Moyle, and KU Docking Fund support to RMB. We also acknowledge University of Kansas Graduate Studies support to CRH, University of Kansas Genome Sequencing Core to both CRH and RMB, USA National Science Foundation (Graduate Research Fellowship award numbers 1540502, 1451148, 0907996 to CRH), systematic panel awards (DEB 1654388, 0743491 to RMB; 1557053 to R. G. Moyle and RMB), and a Doctoral Dissertation Improvement Grant (1701952 to SLT). Additional funding was provided by the Society for the Study of Evolution to CRH. Additionally, SLT was supported by a National Institutes of Health INSPIRE IRACDA postdoctoral fellowship (NIH K12GM093854). We thank the Malagasy authorities for approving research permits (N°298/13/MEF/SG/DGF/DCB.SAP/SCBSE, N°303/14/MEEMF/SG/DGF/DAPT/SCBT, N°329/15/MEEMF/SG/DGF/DAPT/SCBT); specimens were exported under permits: (N°017N-EV01/MG14, N°055N-EA02/MG15, N°041N-EA01/MG16) and the organizations MICET and Centro ValBio for logistics assistance with permits and in the field. The Solomon Islands Forest Department approved field research protocols and export permits (to SLT, RMB, and R. G. Moyle), and the Philippine Department of Environment and Natural Resources (DENR) facilitated Memoranda of Agreement and Gratuitous Permits to Collect Biological Specimens, issued to RMB, 2005–2019. We thank Alana Alexander for providing us with the *Kaloula* UCE data. We also thank David Blackburn of the California Academy of Sciences, Chan Kin Onn of the National University of Singapore, and the Museum of Vertebrate Zoology at Berkeley for tissue samples. This paper constitutes contribution number 947 of the Auburn University Museum of Natural History for PLW. Finally, we thank Arbor BioSciences and their staff scientists for assistance in designing the probes and library preparation, especially Jake Enk and Alison Devault.

AUTHOR CONTRIBUTIONS

The study and experiment were conceived and designed by Carl R. Hutter and Kerry A. Cobb. The probe sets and bioinformatic pipelines were created by Carl R. Hutter. Tissue samples and sequence data were contributed by Carl R. Hutter, Scott L. Travers, and RMB. Laboratory work was performed by Carl R. Hutter, Scott L. Travers, and Perry L. Wood. Data was analysed by Carl R. Hutter, Daniel M.

Portik, and Kerry A. Cobb; website was developed by Kerry A. Cobb. The first draft of the manuscript was written by Carl R. Hutter, where all authors contributed to drafts and approved the final manuscript.

OPEN RESEARCH BADGES

This article has earned an Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at https://osf.io/gvbr5/ and https://github.com/chutter/FrogCap-Sequence-Capture.

DATA AVAILABILITY STATEMENT

A major goal in disseminating these resources is to provide the probe-set described and tested here as a freely available and open access resource. The probe-set and other resources are licensed under a Creative Commons Attribution 3.0 United States (https://creativecommons.org/licenses/by/3.0/us/legalcode). All raw sequencing reads have been made available in the GenBank SRA (BioProject: PRJNA633673) upon acceptance of this manuscript in a peer-reviewed journal. All alignments analysed and materials for replicating analyses are available on the Open Science Framework (Hutter, 2021) (https://osf.io/gvbr5/). Finally, the probe-set and marker files, data analysis pipelines (bioinformatics, variants, and mitochondrial) and the scripts to fully replicate all analyses are available on Carl R. Hutter's GitHub (https://github.com/chutter/FrogCap-Sequence-Capture).

DATA AVAILABILITY STATEMENT

A major goal in disseminating these resources is to provide the probe-set described and tested here as a freely available and open access resource. The probe-set and other resources are licensed under a Creative Commons Attribution 3.0 United States (https://creativecommons.org/licenses/by/3.0/us/legalcode). All raw sequencing reads have been made available in the GenBank SRA (BioProject: PRJNA633673) upon acceptance of this manuscript in a peer-reviewed journal. All alignments analysed and materials for replicating analyses are available on the Open Science Framework (Hutter, 2021) (https://osf.io/gvbr5/). Finally, the probe-set and marker files, data analysis pipelines (bioinformatics, variants, and mitochondrial) and the scripts to fully replicate all analyses are available on Carl R. Hutter's GitHub (https://github.com/chutter/FrogCap-Sequence-Capture).

ORCID

Carl R. Hutter https://orcid.org/0000-0001-6381-6339
Perry L. Wood Jr. https://orcid.org/0000-0003-3767-5274

REFERENCES

Alexander, A. M., Su, Y.-C., Oliveros, C. H., Olson, K. V., Travers, S. L., & Brown, R. M. (2016). Genomic data reveals potential for hybridization, introgression, and incomplete lineage sorting to confound phylogenetic relationships in an adaptive radiation of narrow-mouth frogs. *Evolution*, *71*(2), 475–488. https://doi.org/10.1111/evo.13133

Alexander, R. P., Fang, G., Rozowsky, J., Snyder, M., & Gerstein, M. B. (2010). Annotating non-coding regions of the genome. *Nature Reviews Genetics*, *11*(8), 559–571. https://doi.org/10.1038/nrg2814

AmphibiaWeb (2021). *Information on amphibian biology and conservation*. Retrieved April 8, 2021, from http://www.amphibiaweb.org

Andermann, T., Fernandes, A. M., Olsson, U., Töpel, M., Pfeil, B., Oxelman, B., Aleixo, A., Faircloth, B. C., & Antonelli, A. (2019). Allele phasing greatly improves the phylogenetic utility of ultraconserved elements. *Systematic Biology*, *68*(1), 32–46. https://doi.org/10.1093/sysbio/syy039

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Barrientos, L. S., Streicher, J. W., Miller, E. C., Pie, M. R., Wiens, J. J., & Crawford, A. J. (2021). Phylogeny of terraranan frogs based on 2,665 loci and impacts of missing data on phylogenomic analyses. *Systematics and Biodiversity*, 1–16. https://doi.org/10.1080/14772000.2021.1933249

Beaudry, M. S., Wang, J., Kieran, T. J., Thomas, J., Bayona-Vásquez, N. J., Gao, B., Devault, A., Brunelle, B., Lu, K., Wang, J.-S., Rhodes, O. E., & Glenn, T. C. (2021). Improved microbial community characterization of 16S rRNA via metagenome hybridization capture enrichment. *Frontiers in Microbiology*, *12*, 644662. https://doi.org/10.3389/fmicb.2021.644662

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J., Mattick, J. S., & Haussler, D. (2004). Ultraconserved elements in the human genome. *Science*, *304*(5675), 1321–1325. https://doi.org/10.1126/science.1098119

Bi, K., Vanderpool, D., Singhal, S., Linderoth, T., Moritz, C., & Good, J. M. (2012). Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. *BMC Genomics*, *13*(1), 403. https://doi.org/10.1186/1471-2164-13-403

Blom, M. P. K., Bragg, J. G., Potter, S., & Moritz, C. (2017). Accounting for uncertainty in gene tree estimation: Summary-coalescent species tree inference in a challenging radiation of Australian lizards. *Systematic Biology*, *66*(3), 352–366. https://doi.org/10.1093/sysbio/syw089

Bragg, J. G., Potter, S., Bi, K., & Moritz, C. (2016). Exon capture phylogenomics: Efficacy across scales of divergence. *Molecular Ecology Resources*, *16*(5), 1059–1068. https://doi.org/10.1111/1755-0998.12449

Brandley, M. C., Bragg, J. G., Singhal, S., Chapple, D. G., Jennings, C. K., Lemmon, A. R., Lemmon, E. M., Thompson, M. B., & Moritz, C. (2015). Evaluating the performance of anchored hybrid enrichment at the tips of the tree of life: A phylogenetic analysis of Australian *Eugongylus* group scincid lizards. *BMC Evolutionary Biology*, *15*(62), 1–14. https://doi.org/10.1186/s12862-015-0318-0

Brown, R. M., Siler, C. D., Richards, S. J., Diesmos, A. C., & Cannatella, D. C. (2015). Multilocus phylogeny and a new classification for Southeast Asian and Melanesian forest frogs (family Ceratobatrachidae). *Zoological Journal of the Linnean Society*, *174*(1), 130–168. https://doi.org/10.1111/zoj.12232

Bushnell, B., Rood, J., & Singer, E. (2017). BBMerge – Accurate paired shotgun read merging via overlap. *PLoS One*, *12*(10), e0185056–e185115. https://doi.org/10.1371/journal.pone.0185056

Capella-Gutierrez, S., Silla-Martinez, J. M., & Gabaldon, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Castoe, T. A., de Koning, A. P. J., Kim, H.-M., Gu, W., Noonan, B. P., Naylor, G., Jiang, Z. J., Parkinson, C. L., & Pollock, D. D. (2009). Evidence for an ancient adaptive episode of convergent molecular evolution. *Proceedings of the National Academy of Sciences*, *106*(22), 8986–8991. https://doi.org/10.1073/pnas.0900233106

Chakrabarty, P., Faircloth, B. C., Alda, F., Ludt, W. B., Mcmahan, C. D., Near, T. J., Dornburg, A., Albert, J. S., Arroyave, J., Stiassny, M. L. J., Sorenson, L., & Alfaro, M. E. (2017). Phylogenomic systematics of ostariophysan fishes: Ultraconserved elements support the surprising non-monophyly of Characiformes. *Systematic Biology*, *66*(6), 881–895. https://doi.org/10.1093/sysbio/syx038

Chan, K. O., Hutter, C. R., Wood, P. L., Grismer, L. L., & Brown, R. M. (2020a). Larger, unfiltered datasets are more effective at resolving phylogenetic conflict: Introns, exons, and UCEs resolve ambiguities in Golden-backed frogs (Anura: Ranidae; genus *Hylarana*). *Molecular Phylogenetics and Evolution*, *151*, 106899. https://doi.org/10.1016/j.ympev.2020.106899

Chan, K. O., Hutter, C. R., Wood, P. L., Grismer, L. L., & Brown, R. M. (2020c). Target-capture phylogenomics provide insights on gene and species tree discordances in Old World treefrogs (Anura: Rhacophoridae). *Proceedings of the Royal Society B: Biological Sciences*, *287*(1940), 20202102. https://doi.org/10.1098/rspb.2020.2102

Chan, K. O., Hutter, C. R., Wood, P. L., Grismer, L. L., Das, I., & Brown, R. M. (2020). Gene flow creates a mirage of cryptic species in a Southeast Asian spotted stream frog complex. *Molecular Ecology*, *29*(20), 3970–3987. https://doi.org/10.1111/mec.15603

Chan, K. O., Hutter, C. R., Wood, P. L. Jr, Su, Y.-C., & Brown, R. M. (2021). Gene flow increases phylogenetic structure and inflates cryptic species estimations: a case study on widespread Philippine puddle frogs (*Occidozyga laevis*). *Systematic Biology*, early view, syab034. https://doi.org/10.1093/sysbio/syab034

Chang, Z., Li, G., Liu, J., Zhang, Y. U., Ashby, C., Liu, D., Cramer, C. L., & Huang, X. (2015). Bridger: A new framework for de novo transcriptome assembly using RNA-seq data. *Genome Biology*, *16*(1), 30. https://doi.org/10.1186/s13059-015-0596-2

Charif, D., & Lobry, J. R. (2007). SeqinR 1.0-2: A contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. *Structural Approaches to Sequence Evolution*, *3*, 207–232. Berlin: Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-35306-5_10

Chen, S., Huang, T., Zhou, Y., Han, Y., Xu, M., & Gu, J. (2017). AfterQC: Automatic filtering, trimming, error removing and quality control for fastq data. *BMC Bioinformatics*, *18*(1), 80. https://doi.org/10.1186/s12859-017-1469-3.

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *bioRxiv*, 1–12. https://doi.org/10.1101/274100

Choi, M., Scholl, U. I., Ji, W., Liu, T., Tikhonova, I. R., Zumbo, P., Nayir, A., Bakkaloğlu, A., Özen, S., Sanjad, S., Nelson-Williams, C., Farhi, A., Mane, S., & Lifton, R. P. (2009). Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(45), 19096–19101. https://doi.org/10.1073/pnas.0910672106

Crawford, N. G., Faircloth, B. C., McCormack, J. E., Brumfield, R. T., Winker, K., & Glenn, T. C. (2012). More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. *Biology Letters*, *8*(5), 783–786. https://doi.org/10.1098/rsbl.2012.0331

Dool, S. E., Puechmaille, S. J., Foley, N. M., Allegrini, B., Bastian, A., Mutumi, G. L., Maluleke, T. G., Odendaal, L. J., Teeling, E. C., & Jacobs, D. S. (2016). Nuclear introns outperform mitochondrial DNA in interspecific phylogenetic reconstruction: Lessons from horseshoe bats (Rhinolophidae: *Chiroptera*). *Molecular Phylogenetics and Evolution*, *97*, 196–212. https://doi.org/10.1016/j.ympev.2016.01.003

Edwards, S. V. (2009). Natural selection and phylogenetic analysis. *Proceedings of the National Academy of Sciences*, *106*(22), 8799–8800. https://doi.org/10.1073/pnas.0904103106

Edwards, S. V., Potter, S., Schmitt, C. J., Bragg, J. G., & Moritz, C. (2016). Reticulation, divergence, and the phylogeography–phylogenetics continuum. *Proceedings of the National Academy of Sciences*, *113*(29), 8025–8032. https://doi.org/10.1073/pnas.1601066113

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. *Systematic Biology*, *61*(5), 717–726. https://doi.org/10.1093/sysbio/sys004

Feng, Y. J., Blackburn, D. C., Liang, D., Hillis, D. M., Wake, D. B., Cannatella, D. C., & Zhang, P. (2017). Phylogenomics reveals rapid, simultaneous diversification of three major clades of Gondwanan frogs at the Cretaceous-Paleogene boundary. *Proceedings of the National Academy of Sciences*, *114*(29), E5864–E5870. https://doi.org/10.1073/pnas.1704632114

Frost, D. R., Grant, T., Faivovich, J., Bain, R. H., Haas, A., Haddad, C. F., De sá, R. O., Channing, A., Wilkinson, M., Donnellan, S. C., Raxworthy, C. J., Campbell, J. A., Blotto, B. L., Moler, P., Drewes, R. C., Nussbaum, R. A., Lynch, J. D., Green, D. M., & Wheeler, W. C. (2006). The amphibian tree of life. *Bulletin of the American Museum of Natural History*, *297*, 8–370.

Glaw, F., & Vences, M. (2006). Phylogeny and genus-level classification of mantellid frogs (Amphibia, Anura). *Organisms Diversity & Evolution*, *6*(3), 236–253. https://doi.org/10.1016/j.ode.2005.12.001

Glenn, T. C. (2011). Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*, *11*(5), 759–769. https://doi.org/10.1111/j.1755-0998.2011.03024.x

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. *Nature Biotechnology*, *27*(2), 182–189. https://doi.org/10.1038/nbt.1523

Graybeal, A. (1994). Evaluating the phylogenetic utility of genes: A search for genes informative about deep divergences among vertebrates. *Systematic Biology*, *43*(2), 174–193. https://doi.org/10.1093/sysbio/43.2.174

Guillory, W. X., French, C. M., Twomey, E. M., Chávez, G., Prates, I., von May, R., De la Riva, I., Lötters, S., Reichle, S., Serrano-Rojas, S. J., Whitworth, A., & Brown, J. L. (2020). Phylogenetic relationships and systematics of the Amazonian poison frog genus *Ameerega* using ultraconserved genomic elements. *Molecular Phylogenetics and Evolution*, *142*, 106638. https://doi.org/10.1016/j.ympev.2019.106638

Guillory, W. X., Muell, M. R., Summers, K., & Brown, J. L. (2019). Phylogenomic reconstruction of the Neotropical poison frogs (Dendrobatidae) and their conservation. *Diversity*, *11*(126), 1–14. https://doi.org/10.3390/d11080126

Guo, Y., Long, J., He, J., Li, C.-I., Cai, Q., Shu, X.-O., Zheng, W., & Li, C. (2012). Exome sequencing generates high quality data in non-target regions. *BMC Genomics*, *13*(1), 194. https://doi.org/10.1186/1471-2164-13-194

Halligan, D. L., Eyre-Walker, A., Andolfatto, P., & Keightley, P. D. (2004). Patterns of evolutionary constraints in intronic and intergenic DNA of *Drosophila*. *Genome Research*, *14*(2), 273–279. https://doi.org/10.1101/gr.1329204

Han, K.-L., Braun, E. L., Kimball, R. T., Reddy, S., Bowie, R. C. K., Braun, M. J., Chojnowski, J. L., Hackett, S. J., Harshman, J., Huddleston, C. J., Marks, B. D., Miglia, K. J., Moore, W. S., Sheldon, F. H., Steadman, D. W., Witt, C. C., & Yuri, T. (2011). Are transposable element insertions homoplasy free?: An examination using the avian tree of life. *Systematic Biology*, *60*(3), 375–386. https://doi.org/10.1093/sysbio/syq100

Hancock-Hanser, B. L., Frey, A., Leslie, M. S., Dutton, P. H., Archer, F. I., & Morin, P. A. (2013). Targeted multiplex next-generation sequencing: advances in techniques of mitochondrial and nuclear DNA sequencing for population genomics. *Molecular Ecology Resources*, *13*, 254–268. https://doi.org/10.1111/1755-0998.12059

Hedge, J., & Wilson, D. J. (2014). Bacterial phylogenetic reconstruction from whole genomes is robust to recombination but demographic inference is not. *MBio*, *5*(6), e02158–14. https://doi.org/10.1128/mBio.02158-14

Hedtke, S. M., Morgan, M. J., Cannatella, D. C., & Hillis, D. M. (2013). Targeted enrichment: Maximizing orthologous gene comparisons across deep evolutionary time. *PLoS One*, *8*(7), e67908. https://doi.org/10.1371/journal.pone.0067908

Heinicke, M. P., Lemmon, A. R., Lemmon, E. M., McGrath, K., & Hedges, S. B. (2018). Phylogenomic support for evolutionary relationships of New World direct-developing frogs (Anura: Terraranae). *Molecular Phylogenetics and Evolution*, *118*, 145–155. https://doi.org/10.1016/j.ympev.2017.09.021

Hellsten, U., Harland, R. M., Gilchrist, M. J., Hendrix, D., Jurka, J., Kapitonov, V., Ovcharenko, I., Putnam, N. H., Shu, S., Taher, L., Blitz, I. L., Blumberg, B., Dichmann, D. S., Dubchak, I., Amaya, E., Detter, J. C., Fletcher, R., Gerhard, D. S., Goodstein, D., … Rokhsar, D. S. (2010). The genome of the western clawed frog *Xenopus tropicalis*. *Science*, *328*(5978), 633–636. https://doi.org/10.1126/science.1183670

Hime, P. M., Lemmon, A. R., Lemmon, E. C. M., Prendini, E., Brown, J. M., Thomson, R. C., Kratovil, J. D., Noonan, B. P., Pyron, R. A., Peloso, P. L. V., Kortyna, M. L., Keogh, J. S., Donnellan, S. C., Mueller, R. L., Raxworthy, C. J., Kunte, K., Ron, S. R., Das, S., Gaitonde, N., … Weisrock, D. W. (2021). Phylogenomics reveals ancient gene tree discordance in the amphibian tree of life. *Systematic Biology*, *70*, 49–66. https://doi.org/10.1093/sysbio/syaa034

Hosner, P. A., Braun, E. L., & Kimball, R. T. (2016). Rapid and recent diversification of curassows, guans, and chachalacas (Galliformes: Cracidae) out of Mesoamerica: Phylogeny inferred from mitochondrial, intron, and ultraconserved element sequences. *Molecular Phylogenetics and Evolution*, *102*, 320–330. https://doi.org/10.1016/j.ympev.2016.06.006

Hugall, A. F., O'Hara, T. D., Hunjan, S., Nilsen, R., & Moussalli, A. (2016). An exon-capture system for the entire class Ophiuroidea. *Molecular Biology and Evolution*, *33*, 281–294. https://doi.org/10.1093/molbev/msv216

Hutter, C. R. (2021). *FrogCap: a modular sequence capture probe set for phylogenomics and population genetics for all frogs, assessed across multiple phylogenetic scales*. Retrieved from osf.io/gvbr5

Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., Ho, S. Y. W., Faircloth, B. C., Nabholz, B., Howard, J. T., Suh, A., Weber, C. C., da Fonseca, R. R., Li, J., Zhang, F., Li, H., Zhou, L., Narula, N., Liu, L., … Zhang, G. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, *346*(6215), 1320–1331. https://doi.org/10.1126/science.1253451

Jones, M. R., & Good, J. M. (2016). Targeted capture in evolutionary and ecological genomics. *Molecular Ecology*, *25*, 185–202. https://doi.org/10.1111/mec.13304

Karin, B. R., Gamble, T., & Jackman, T. R. (2020). Optimizing phylogenomics with rapidly evolving long exons: Comparison with anchored hybrid enrichment and ultraconserved elements. *Molecular Biology and Evolution*, *37*(3), 904–922. https://doi.org/10.1093/molbev/msz263

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Katzman, S., Kern, A. D., Bejerano, G., Fewell, G., Fulton, L., Wilson, R. K., Salama, S. R., & Haussler, D. (2007). Human genome ultraconserved elements are ultraselected. *Science*, *317*(5840), 915. https://doi.org/10.1126/science.1142430

Kent, W. J. (2002). BLAT—the BLAST-like alignment tool. *Genome Research*, *12*(4), 656–664. https://doi.org/10.1101/gr.229202

Kircher, M., & Kelso, J. (2010). High-throughput DNA sequencing – concepts and limitations. *BioEssays*, *32*, 524–536. https://doi.org/10.1002/bies.200900181

Lanier, H. C., & Knowles, L. L. (2012). Is recombination a problem for species-tree analyses? *Systematic Biology*, *61*(4), 691–701. https://doi.org/10.1093/sysbio/syr128

Laurence, M., Hatzis, C., & Brash, D. E. (2014). Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. *PLoS One*, *9*(5), e97876–e97878. https://doi.org/10.1371/journal.pone.0097876

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T., & Carey, V. J. (2013). Software for computing and annotating genomic ranges. *PLOS Computational Biology*, *9*(8), e1003118. https://doi.org/10.1371/journal.pcbi.1003118

Leaché, A. D., & Oaks, J. R. (2017). The utility of single nucleotide polymorphism (SNP) data in phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, *48*, 69–84. https://doi.org/10.1146/annurev-ecolsys-110316-022645

Lemmon, A. R., Emme, S. A., & Lemmon, E. M. (2012). Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*, *61*(5), 727–744. https://doi.org/10.1093/sysbio/sys049

Lemmon, E. M., & Lemmon, A. R. (2013). High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, *44*, 99–121. https://doi.org/10.1146/annurev-ecolsys-110512-135822

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22*(13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Li, Y.-L., Weng, J.-C., Hsiao, C.-C., Chou, M.-T., Tseng, C.-W., & Hung, J.-H. (2015). PEAT: An intelligent and efficient paired-end sequencing adapter trimming algorithm. *BMC Bioinformatics*, *16*(1), S2. https://doi.org/10.1186/1471-2105-16-S1-S2

Liu, L., Yu, L., Pearl, D. K., & Edwards, S. V. (2009). Estimating species phylogenies using coalescence times among sequences. *Systematic Biology*, *58*(5), 468–477. https://doi.org/10.1093/sysbio/syp031

Maddison, W. P. (1997). Gene trees in species trees. *Systematic Biology*, *46*(3), 523–536. https://doi.org/10.1093/sysbio/46.3.523

McCartney-Melstad, E., Mount, G. G., & Shaffer, H. B. (2016). Exon capture optimization in amphibians with large genomes. *Molecular Ecology Resources*, *16*, 1084–1094. https://doi.org/10.1111/1755-0998.12538

McCormack, J. E., Hird, S. M., Zellmer, A. J., Carstens, B. C., & Brumfield, R. T. (2013). Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, *66*, 526–538. https://doi.org/10.1016/j.ympev.2011.12.007

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110

Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., & Johnson, E. A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, *17*, 240–248. https://doi.org/10.1101/gr.5681207

Mirarab, S., Bayzid, M. S., Boussau, B., & Warnow, T. (2014). Statistical binning enables an accurate coalescent-based estimation of the avian tree. *Science*, *346*(6215), 1250463. https://doi.org/10.1126/science.1250463

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, *5*(7), 621–628. https://doi.org/10.1038/nmeth.1226

Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Peloso, P. L. V., Frost, D. R., Richards, S. J., Rodrigues, M. T., Donnellan, S., Matsui, M., Raxworthy, C. J., Biju, S. D., Lemmon, E. M., Lemmon, A. R., & Wheeler, W. C. (2016). The impact of anchored phylogenomics and taxon sampling on phylogenetic inference in narrow-mouthed frogs (Anura, Microhylidae). *Cladistics*, *32*(2), 113–140. https://doi.org/10.1111/cla.12118

Pie, M. R., Faircloth, B. C., Ribeiro, L. F., Bornschein, M. R., & McCormack, J. E. (2018). Phylogenomics of montane frogs of the Brazilian Atlantic Forest is consistent with isolation in sky islands followed by climatic stability. *Biological Journal of the Linnean Society*, *125*(1), 72–82. https://doi.org/10.1093/biolinnean/bly093

Portik, D. M., Smith, L. L., & Bi, K. (2016). An evaluation of transcriptome-based exon capture for frog phylogenomics across multiple scales of divergence (Class: Amphibia, Order: Anura). *Molecular Ecology Resources*, *16*(5), 1069–1083. https://doi.org/10.1111/1755-0998.12541

Portik, D. M., & Wiens, J. J. (2021). Do alignment and trimming methods matter for phylogenomic (UCE) analyses? *Systematic Biology*, *70*(3), 440–462. https://doi.org/10.1093/sysbio/syaa064

Prum, R. O., Berv, J. S., Dornburg, A., Field, D. J., Townsend, J. P., Lemmon, E. M., & Lemmon, A. R. (2015). A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, *526*(7574), 569–573. https://doi.org/10.1038/nature15697

Pyron, R. A., & Wiens, J. J. (2011). A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. *Molecular Phylogenetics and Evolution*, *61*(2), 543–583. https://doi.org/10.1016/j.ympev.2011.06.012

Rasolonjatovo, S. M., Scherz, M. D., Hutter, C. R., Glaw, F., Rakotoarison, A., Razafindraibe, J. H., Goodman, S. M., Raselimanana, A. P., & Vences, M. (2020). Sympatric lineages in the Mantidactylus ambreensis complex of Malagasy frogs originated allopatrically rather than by in-situ speciation. *Molecular Phylogenetics and Evolution*, *144*, 106700. https://doi.org/10.1016/j.ympev.2019.106700

Reddy, S., Kimball, R. T., Pandey, A., Hosner, P. A., Braun, M. J., Hackett, S. J., Han, K.-L., Harshman, J., Huddleston, C. J., Kingston, S., Marks, B. D., Miglia, K. J., Moore, W. S., Sheldon, F. H., Witt, C. C., Yuri, T., & Braun, E. L. (2017). Why do phylogenomic data sets yield conflicting trees? Data type influences the avian tree of life more than taxon sampling. *Systematic Biology*, *66*(5), 857–879. https://doi.org/10.1093/sysbio/syx041

Reilly, S. B., Stubbs, A. L., Karin, B. R., Bi, K., Arida, E., Iskandar, D. T., & McGuire, J. A. (2019). Leap-frog dispersal and mitochondrial introgression: Phylogenomics and biogeography of Limnonectes fanged frogs in the Lesser Sundas Archipelago of Wallacea. *Journal of Biogeography*, *46*(4), 757–769. https://doi.org/10.1111/jbi.13526

Richards, E. J., Brown, J. M., Barley, A. J., Chong, R. A., & Thomson, R. C. (2018). Variation across mitochondrial gene trees provides evidence for systematic error: How much gene tree variation is biological? *Systematic Biology*, *67*(5), 847–860. https://doi.org/10.1093/sysbio/syy013

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, *4*(17), e2584–e2622. https://doi.org/10.7717/peerj.2584

Rohland, N., & Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, *22*(5), 939–946. https://doi.org/10.1101/gr.128124.111

Safonova, Y., Bankevich, A., & Pevzner, P. A. (2015). dipSPAdes: Assembler for highly polymorphic diploid genomes. *Journal of Computational Biology*, *22*(6), 528–545. https://doi.org/10.1089/cmb.2014.0153

Simmons, M. P., & Ochoterena, H. (2000). Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology*, *49*(2), 369–381. https://doi.org/10.1093/sysbio/49.2.369

Singhal, S., Grundler, M., Colli, G., & Rabosky, D. L. (2017). Squamate Conserved Loci (SqCL): A unified set of conserved loci for phylogenomics and population genetics of squamate reptiles. *Molecular Ecology Resources*, *17*(6), e12–e24. https://doi.org/10.1111/1755-0998.12681

Smith, B. T., Harvey, M. G., Faircloth, B. C., Glenn, T. C., & Brumfield, R. T. (2014). Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Systematic Biology*, *63*(1), 83–95. https://doi.org/10.1093/sysbio/syt061

Springer, M. S., & Gatesy, J. (2018). Delimiting coalescence genes (C-Genes) in phylogenomic data sets. *Genes*, *9*(123), 1–19. https://doi.org/10.3390/genes9030123

Stephen, S., Pheasant, M., Makunin, I. V., & Mattick, J. S. (2008). Large-scale appearance of ultraconserved elements in tetrapod genomes and slowdown of the molecular Clock. *Molecular Biology and Evolution*, *25*(2), 402–408. https://doi.org/10.1093/molbev/msm268

Streicher, J. W., Loader, S. P., Varela-Jaramillo, A., Montoya, P., & de Sá, R. O. (2020). Analysis of ultraconserved elements supports African origins of narrow-mouthed frogs. *Molecular Phylogenetics and Evolution*, *146*, 106771. https://doi.org/10.1016/j.ympev.2020.106771

Streicher, J. W., Miller, E. C., Guerrero, P. C., Correa, C., Ortiz, J. C., Crawford, A. J., Pie, M. R., & Wiens, J. J. (2018). Evaluating methods for phylogenomic analyses, and a new phylogeny for a major frog clade (Hyloidea) based on 2214 loci. *Molecular Phylogenetics and Evolution*, *119*, 128–143. https://doi.org/10.1016/j.ympev.2017.10.013

Sulonen, A.-M., Ellonen, P., Almusa, H., Lepistö, M., Eldfors, S., Hannula, S., Miettinen, T., Tyynismaa, H., Salo, P., Heckman, C., Joensuu, H., Raivio, T., Suomalainen, A., & Saarela, J. (2011). Comparison of solution-based exome capture methods for next generation sequencing. *Genome Biology*, *12*(9), R94. https://doi.org/10.1186/gb-2011-12-9-r94

Sun, Y.-B., Xiong, Z.-J., Xiang, X.-Y., Liu, S.-P., Zhou, W.-W., Tu, X.-L., & Zhang, Y. (2015). Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. *Proceedings of the National Academy of Sciences*, *112*(11), E1257–E1262. https://doi.org/10.1073/pnas.1501764112

Taylor, D. J., & Piel, W. H. (2004). An assessment of accuracy, error, and conflict with support values from genome-scale phylogenetic data. *Molecular Biology and Evolution*, *21*(8), 1534–1537. https://doi.org/10.1093/molbev/msh156

Tewhey, R., Nakano, M., Wang, X., Pabón-Peña, C., Novak, B., Giuffre, A., Lin, E., Happe, S., Roberts, D. N., LeProust, E. M., Topol, E. J., Harismendy, O., & Frazer, K. A. (2009). Enrichment of sequencing targets from the human genome by solution hybridization. *Genome Biology*, *10*(10), R116. https://doi.org/10.1186/gb-2009-10-10-r116

Townsend, J. (2007). Profiling phylogenetic informativeness. *The Auk*, *56*(2), 222–231. https://doi.org/10.1080/10635150701311362

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, *11*(1100), 11.10.1–11.10.33. https://doi.org/10.1002/0471250953.bi1110s43

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*(1), 57–63. https://doi.org/10.1038/nrg2484

Xi, Z., Liu, L., & Davis, C. C. (2015). Genes with minimal phylogenetic information are problematic for coalescent analyses when gene tree estimation is biased. *Molecular Phylogenetics and Evolution*, *92*, 63–71. https://doi.org/10.1016/j.ympev.2015.06.009

Yuan, Z.-Y., Zhang, B.-L., Raxworthy, C. J., Weisrock, D. W., Hime, P. M., Jin, J.-Q., Lemmon, E. M., Lemmon, A. R., Holland, S. D., Kortyna, M. L., Zhou, W.-W., Peng, M.-S., Che, J., & Prendini, E. (2019). Natatanuran frogs used the Indian Plate to step-stone disperse and radiate across the Indian Ocean. *National Science Review*, *6*(1), 10–14. https://doi.org/10.1093/nsr/nwy092

Zarza, E., Faircloth, B. C., Tsai, W. L. E., Bryson, R. W. Jr, Klicka, J., & McCormack, J. E. (2016). Hidden histories of gene flow in highland birds revealed with genomic markers. *Molecular Ecology*, *25*(20), 5144–5157. https://doi.org/10.1111/mec.13813

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Hutter, C. R., Cobb, K. A., Portik, D. M., Travers, S. L., Wood, P. L. Jr., & Brown, R. M. (2022). FrogCap: A modular sequence capture probe-set for phylogenomics and population genetics for all frogs, assessed across multiple phylogenetic scales. *Molecular Ecology Resources*, 22, 1100–1119. https://doi.org/10.1111/1755-0998.13517