Reports

Capturing protein-coding genes across highly divergent species

Chenhong Li[1,3], Michael Hofreiter[2], Nicolas Straube[3], Shannon Corrigan[3], and Gavin J.P. Naylor[3]
[1]Key Laboratory of Exploration and Utilization of Aquatic Genetic Resources, Shanghai Ocean University, Ministry of Education, Shanghai, China, [2]Department of Biology, The University of York, Heslington, UK, and [3]Hollings Marine Laboratory, College of Charleston, Charleston, SC, USA

BioTechniques 54:321-326 (June 2013) doi 10.2144/000114039
Keywords: comparative biology; cross-species capture; DNA hybridization capture; next-generation sequencing; protein-coding genes; phylogenetics; target enrichment

Supplementary material for this article is available at www.BioTechniques.com/article/114039.

DNA hybridization capture combined with next generation sequencing can be used to determine the sequences of hundreds of target genes across hundreds of individuals in a single experiment. However, the approach has thus far only been successfully applied to capture targets that are highly similar in sequence to the bait molecules. Here we introduce modifications that extend the reach of the method to allow efficient capture of highly divergent homologous target sequences using a single set of baits. These modifications have important implications for comparative biology.

Next generation sequencing is revolutionizing our understanding of the links between genotype and phenotype at scales that were unimaginable a few years ago. Genome-wide association studies are now routinely conducted to explore the genomic basis for phenotypic variation (1) or disease predisposition (2), as well as to identify the genomic elements that underlie adaptive evolutionary responses (3).

Although such genome level association studies are powerful, they are labor-intensive to execute, as they require not only sequencing of entire genomes, but also that the collected genomic data be carefully parsed, assembled, and accurately annotated. This is most easily done when closely related and well-annotated reference genomes are available for comparison. However, genome assembly and annotation become substantially more difficult with increasing evolutionary distance to the reference genome. Indeed, it is often easier to assemble the genomes of divergent species de novo, than it is to compare them to a reference that differs in genome organization. This said, most studies do not require whole genome comparisons. Comparative medical and physiological studies, for example, usually center on a specific set of genes relevant to a pathway of interest (4). In the same vein, most molecular evolutionary and phylogenetic studies focus on the comparison of small subsets of homologous genes.

Unfortunately, no technology yet exists that allows researchers to efficiently explore genetic variation across evolutionarily divergent organisms for large sets of pre-specified target genes, such as those that determine pelage color in vertebrates (5), or those involved in particular physiological adaptations (6,7). PCR amplification, the main and for a long time only technology for targeting specific DNA regions for sequencing, is generally too labor intensive, costly, and inconsistent to generate such multigene data sets. In principle, DNA hybridization capture technology, hereafter referred to as gene capture (8,9), should be up to the task as it allows for hundreds of pre-specified genes to be targeted and isolated for sequencing in a single experiment. Indeed, gene capture techniques do work well when baits are used to interrogate libraries with limited sequence divergence (10–13), such as those for the same species or for closely related species, or when baits are designed to target ultraconserved elements among species that otherwise have highly divergent genomic sequences (14,15). Unfortunately, existing gene capture protocols do not work well when targeting genes whose sequences are highly divergent from those of the baits that are used to capture them, although we know from first principles and experience with Southern blotting (16), that they should.

Here we describe a gene capture method that is effective for capturing a pre-specified set of protein coding genes across species that have been evolving independently for hundreds of millions of years, using a single bait array. We have achieved this by tuning the stringency of the hybridization and washing steps of the procedure to optimize retention of target versus non-target DNA fragments, and by conducting two rounds of gene capture whereby the captured products from the first round are used as templates to augment a second round of gene capture (see detailed Supplementary Protocol).

Method summary:

Modified DNA hybridization conditions allow gene capture across widely divergent homologous sequences.

Materials and methods

Proof of concept

To demonstrate the efficacy of gene capture across highly divergent taxa we selected a set of target genes that would be (*i*) present in all taxa tested and (*ii*) unambiguously identifiable (unique) within each genome. We targeted coding DNA sequences (CDS) from single-copy protein-coding genes that are shared across gnathostome vertebrates. We compared the published and annotated genomes of human (*Homo sapiens*), chicken (*Gallus gallus*), western clawed toad (*Xenopus tropicalis*), green anole (*Anolis carolinensis*), zebrafish (*Danio rerio*), and elephant shark (*Callorhinchus milii*), and identified the putatively orthologous genes within each respective genome. A within-species BLAST search (http://blast.ncbi.nlm.nih.gov/Blast.cgi) (17) was used to identify gene regions determined to be single-copy according to a 60% dissimilarity criterion (i.e., there were no genomic regions present that were more than 40% similar to the potential target according to the within-species BLAST search criterion). We then used among-species BLAST searches to verify that the identified targets occurred in all six vertebrates. The EvolMarkers software for selecting such sequence sets for any combination of species is freely available at http://bioinformatics.unl.edu/cli/evolmarkers/ (18). We identified a total of 1449 candidate single-copy CDS markers that were shared across the 6 vertebrate species using this approach. The size of the target CDS varied from 151 bp to 2390 bp.

Bait development

We developed separate custom biotinylated RNA bait libraries based on the identified CDS for the *H. sapiens, G. gallus, X. tropicalis, A. carolinensis, D. rerio,* and *C. milii* genomes using the MYBaits target enrichment system (MYcroarray, Ann Arbor, Michigan). Each resulting bait library comprised a pooled series of 120 bp baits designed for each target gene. We included a 60 bp overlap between baits, allowing for 2× complete coverage of each target gene (2 × tiling). If the length of the target gene was less than 120 bp, the sequence was extended in length to 120 bp by adding thymine (T) nucleotides.

Experimental design

We carried out two experiments. First, in an effort to evaluate the effectiveness of the method across broadly different classes of gnathostome vertebrates, we used baits to capture the selected set of 1449 target single-copy genes from (*i*) genomic libraries of the species from which the baits were derived (as a positive (+ve) control) and (*ii*) distantly related species belonging to the same vertebrate class as the species from which the bait sequence was derived. Thus, for fish we used baits designed from the *D. rerio* genome to capture target DNA sequences from *D. rerio* (+ve control) and the stickleback (*Gasterosteus aculeatus*); for amphibians we used baits from the *X. tropicalis* genome to capture targets in *X. tropicalis* (+ve control) and the axolotl (*Ambystoma mexicanum*); for reptiles we used baits from the *A. carolinensis* genome to capture targets in *A. carolinensis* (+ve control) and the painted turtle (*Chrysemys picta*); for birds we used baits from the *G. gallus* genome to capture targets in *G. gallus* (+ve control) and the zebra finch (*Taeniopygia guttata*); and for mammals we used baits from the *H. sapiens* genome to capture target genes in *H. sapiens* (+ve control) and the gray short-tailed opossum (*Monodelphis domestica*). The similarities of the target CDS ranged from 89% to 61% among the 6 vertebrates (ultraconserved element approaches target highly conserved sequences with similarities typically greater than 90% (14,15)).

We compared the effects of two different hybridization and washing schemes on the capture results. Standard conditions comprised 65°C for both hybridization and washing temperature, following the manual for the MYBaits Target Enrichment System (Mycroarray, Ann Arbor, Michigan). The relaxed conditions comprised hybridization under a touchdown gene capture scheme similar to Mason et al. (10) but with a much lower final temperature: 65°C for 11 h, followed by 60°C for 11 h, 55°C for 11 h, and 50°C for 11 h, using 45°C instead of 65°C for the last three washing steps (see detailed protocol). We also tested whether capturing twice (i.e., using the captured products after re-amplification as templates to perform a second round of capture) increases the number of captured target genes.

Secondly, in an effort to explore the effectiveness of the method within a class of gnathostomes, we applied the optimized protocol to a diverse suite of chondrichthyan fishes (sharks, skates, rays, and chimaeras), the group that represents the most basal gnathostome class and whose origins date back to the Devonian period, more than 400 million years ago (19). We used baits based on the elephant shark (*C. milii*) genome to capture the 1449 target genes from *C. milii* (+ve control), five skates and rays (*Aetobatos narinari*, *Leucoraja erinacea*, *Neotrygon kuhlii*, *Rhinobatos schlegelii*, *Torpedo formosa*), and seven sharks (*Carcharhinus amblyrhynchos*, *Chlamydoselachus anguineus*, *Etmopterus joungi*, *Heterodontus portusjacksoni*, *Isurus oxyrinchus*, *Orectolobus halei*, *Squatina nebulosa*).

For detailed experimental design information, please refer to the protocols for library preparation using the with-bead method and gene capture that are provided as Supplementary Material.

Sequencing and analysis

We pooled 20 indexed samples in equimolar ratios after gene capture. We then measured the pooled product using the CFX Connect Real-Time PCR system (Bio-Rad, Hercules, CA) and used 600 µL of 8 pM sample for

Table 1. The number of the 1449 target CDS captured using different protocols, for 10 vertebrates.

Bait sp.	*D. rerio*		*X. tropicalis*		*A. carolinensis*		*G. gallus*		*H. sapiens*	
Target sp.	*D. rerio*	*G. aculeatus*	*X. tropicalis*	*A. mexicanum*	*A. carolinensis*	*C. picta*	*G. gallus*	*T. guttata*	*H. sapiens*	*M. domestica*
Genome size (pg)	1.68–2.28	0.58–0.70	1.78	21.85–48	1.61–3.06	2.57–3.22	1.25	1.25	3.5	3.5
Time of divergence from bait species (MYA)	~ 0	~ 254.4	~ 0	~ 298.6	~ 0	~ 244.2	~ 0	~ 106.4	~ 0	~ 163.9
[1]Standard capture	1216 (84%)	67 (4.6%)	1125 (78%)	0	958 (66%)	14 (1%)	1259 (87%)	554 (38%)	1124 (78%)	132 (9%)
[2]Relaxed 1st capture	1349 (93%)	329 (23%)	1412 (97%)	53 (4%)	1365 (94%)	377 (26%)	1420 (98%)	1057 (73%)	1297 (90%)	438 (30%)
[3]Relaxed 2nd capture	1387 (96%)	643 (44%)	1428 (99%)	225 (16%)	1374 (95%)	777 (54%)	1413 (98%)	1159 (80%)	1355 (94%)	1027 (71%)

[1]Standard capture: the standard protocol using 65°C for the hybridization and the second washing step. [2]Relaxed first capture: relaxed hybridization conditions, i.e., touchdown hybridization scheme of 65°C for 11 h, followed by 60°C for 11 h, followed by 55°C for 11 h,completed by 50°C for 11 h, and using 45°C instead of 65°C for the final 3 washing steps (see detailed protocol). [3]Relaxed second capture: products from the first capture were used as templates to perform another round of capture. Divergence times were calculated from "The TimeTree of Life" (29). The genome size data were retrieved from the Animal Genome Size Database (www.genomesize.com).

paired-end 150 bp sequencing on a MiSeq sequencer (Illumina, Inc, San Diego, CA). Sequence reads associated with each sample were identified by their respective indices. The adapter sequences and sites with lower qualities were trimmed using the Cutadapt application (version 1.1; http://code.google.com/p/cutadapt/)(20). Contigs were assembled de novo for each species using ABySS (version 1.3.4; www.bcgsc.ca/platform/bioinfo/software/abyss) (21). A custom Perl script (Supplementary Material) was used to retrieve the sequences for each gene for each species studied from the assembled contigs. The aligned sequences for each gene were checked by eye to ensure accuracy using Geneious Pro v5.6.2 (Biomatters, Auckland, New Zealand; available at www.geneious.com/). To ensure that only orthologous sequences were used in our final analyses, we only considered cases where the target returned a single hit. We were able to do this because our bait arrays were explicitly designed to target single copy genes in the 6 model organisms (based on a conservative 60% dissimilarity criterion used in the initial selection of target sequences). As a further test of orthology, we used a Hidden Markov Model (HMM), trained on a curated set of alignments of the 1449 single copy targets across the 6 model organisms (22) to assign captured sequences to their corresponding orthologous targets. In all cases, more than 90% of the captured sequences were assigned as orthologs by the HMM. We excluded all genes whose orthology was questionable from further analysis. We caution that paralogs resulting from gene duplication and subsequent asymmetric loss cannot always be ruled out with our approach. The average identity between the captured sequences and the baits was calculated using custom Perl scripts (Supplementary Material).

Results and discussion

Comparison across different classes of gnathostome vertebrates

The results we obtained varied depending on hybridization and washing conditions (Table 1). Under standard conditions, very few target sequences were successfully retrieved, except for the positive controls, where both bait and target library were derived from the same species. The only exception to this was the result obtained for birds where 554 target sequences were captured for *T. guttata* (zebra finch) using *G. gallus* (chicken) baits. We speculate that this is a consequence of the unusually high level of sequence similarity seen among divergent species of birds relative to their non-avian vertebrate counterparts (23). By contrast, under the relaxed hybridization conditions, there was an eight-fold increase in the number of target sequences captured relative to experiments executed under standard conditions. The difference in the effectiveness of gene capture was especially obvious where target species were highly divergent from bait species (Table 1).

Performing 2 rounds of gene capture further increased the number of targets captured relative to the initial round of capture by 68%, on average. Once again, the improvement was most conspicuous where the bait and target species were highly divergent (see Table 1). The effectiveness of cross-species gene capture with the optimized protocol ranged from a minimum of 225 (for the amphibians) to a maximum of 1159 (between chicken and zebra finch) of a possible 1449 target CDS. The lowest number of target genes successfully captured (225) resulted when baits designed from the *X. tropicalis* genome were used to capture corresponding genes in the axolotl, *A. mexicanum*. We hypothesize that this result may be due to compromising effects associated with repeats in the unusually large (~32×10^9 bp) (24) genome of the axolotl. The protocol we used for target enrichment used human *cot-1* DNA in an effort to block and therefore mitigate the adverse effect of repetitive DNA during hybridization. We suspect that tailored *cot-1* DNA derived from *A. mexicanum* may be necessary to effectively block repeat elements impeding capture for the axolotl.

We investigated whether factors other than bait-target divergence affected the efficacy of gene capture. We compared the GC content, target sequence length, and chromosomal position between the set of targets that were successfully captured, and those that failed capture. This was carried out for the six positive controls (*H. sapiens, G. gallus, X. tropicalis, A. carolinensis, D. rerio,* and *C. milii*,) where the bait and target sequences are identical, allowing us to rule out any confounding effects that might be caused by sequence divergence. On average, the few targets that were not captured had higher GC contents (56%) and shorter lengths (235 bp) than those that were successfully captured (GC = 49%, length = 303 bp). However, there was overlap in both GC content and target length between the captured and non-captured targets suggesting that other influences may be involved when targets are not captured. Chromosomal position did not seem to have any effect on capture success.

In summary, the efficacy of gene capture was improved by incorporating both touchdown gene capture and conducting a second round of capture, but there was considerable variation in efficacy across the five classes of gnathostome vertebrates tested. We hypothesize that this was due to differences in rates of molecular evolution among the pairs of vertebrates, the presence of genomic anomalies such as repeats that are known to interfere with gene capture (25), or secondary structural features that inhibited hybridization to the baits (26).

Table 2. The number of the 1449 target CDS captured for 13 chondrichthyans using baits based on the *C. milii* genome and the optimized capture protocol.

Target species	No. of target CDS captured	No. of CDS kept for final analysis
Callorhinchus milii	1449	1242
Aetobatus narinari	1309	1182
Leucoraja erinacea	1004	1232*
Neotrygon kuhlii	1284	1163
Rhinobatos schlegelii	1351	1212
Torpedo formosa	1036	928
Carcharhinus amblyrhynchos	1283	1150
Chlamydoselachus anguineus	1217	1101
Etmopterus joungi	1082	976
Heterodontus portusjacksoni	1294	1167
Isurus oxyrinchus	1251	1136
Orectolobus halei	1206	1087
Squatina nebulosa	1192	1080

*Additional target sequences were retrieved from genome data available in GenBank (http://www.ncbi.nlm.nih.gov).

Comparison within a class of vertebrates: chondrichthyan fishes

A total of 13 hybridization reactions were carried out [elephant shark (*C. milii*), five skates and rays (*Aetobatus narinari*, *Leucoraja erinacea*, *Neotrygon kuhlii*, *Rhinobatos schlegelii*, *Torpedo formosa*), and seven sharks (*Carcharhinus amblyrhynchos*, *Chlamydoselachus anguineus*, *Etmopterus joungi*, *Heterodontus portusjacksoni*, *Isurus*

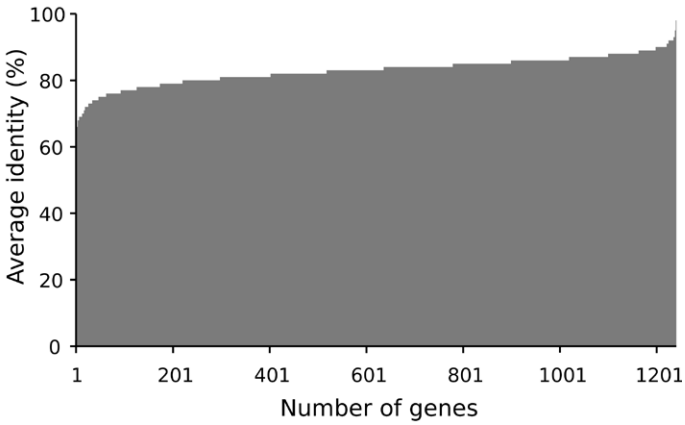

Figure 1. Average identity of the 1242 successfully captured gene sequences when compared with the bait sequences of *C. milii*.

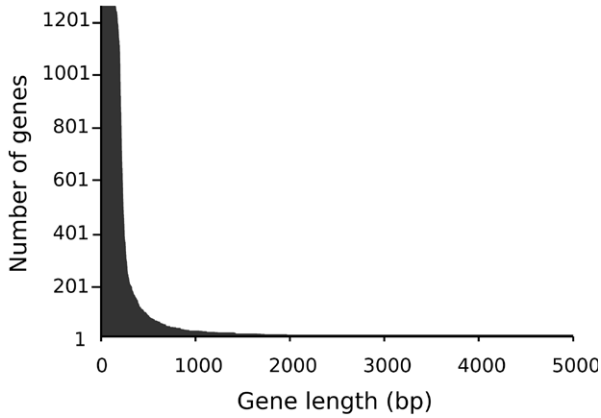

Figure 2. Length distribution of the captured genes for chondrichthyan fishes.

oxyrinchus, *Orectolobus halei*, *Squatina nebulosa*)]. Touchdown gene capture and the double capture protocol were deployed as in the first set of experiments across classes of gnathostome vertebrates. For the cross-species captures, we were able to obtain full sequences for 1004 of the 1449 target sequences in the worst case, and 1351 of 1449 target sequences in the best case (Table 2). We obtained 1449 of 1449 sequences for the positive control (*C. milii* baits tested against a *C. milii* library). These results show greater homogeneity in the efficacy of capture than was evident in the survey across vertebrate classes, confirming the expectation that the methods would be more consistent when applied to a denser taxonomic sample of more closely related lineages. A small subset of genes that were either not captured for the majority of taxa, or showed some evidence of potential paralogy due to local gene duplication were excluded from our final data set.

Overall, more than 90% of the captured sequences were assigned as orthologs by HaMSTR (22). The average identity between baits and successfully captured target sequences ranged from 61% to 98% (Figure 1). In total, we obtained 338,822 orthologous base pairs (of 418,475 initially targeted) across the 13 chondrichthyan orders (i.e., 81% of the original target set) with a maximum sequence divergence between bait and target of 39%. The final chondrichthyan data set included 1242 successfully captured putatively orthologous CDS that ranged in length from 112 to 5091 bp, with 51 CDS being more than 600 bp in length (Figure 2).

Capturing homologous genes across species is not a new idea. However, previous approaches have either (*i*) focused on bait-target divergences that are much shallower than those used in the current study (12,13), or (*ii*) targeted highly conserved regions in an effort to extract useful sequence information from the associated flanking regions that are generally more variable (14,15,27,28). The method described here is distinct from both of these approaches. It is designed to explicitly target known orthologous protein-coding genes across a range of divergences (up to 39% bait-target dissimilarity) using a single set of probes. This has been achieved by maximizing retention of target material by significantly lowering the hybridization temperature (e.g., our touchdown hybridization temperature ended at 50°C in comparison to 60°C used by Mason et al.) (10); by lowering the temperature of the second wash to 45°C compared to 65°C under the standard protocol; and by deploying 2 rounds of enrichment.

By way of comparison, the ultraconserved element (14,27,28) and anchored enrichment methods (15) are strategies to determine sequences in the variable regions that flank known conserved genomic elements. A significant proportion of the sequences that occur in these flanking regions have been found to be noncoding, or of unassigned function. The rate of evolutionary change per site generally increases exponentially across these flanking regions with increasing distance from the conserved ultraconserved element core. Such patterns of rate variation have been reported as an asset by advocates of ultraconserved element methods (15) because they allow for the collection of sequences exhibiting a range of evolutionary rates, which may be helpful for identifying markers relevant to particular phylogenetic questions. However, these same patterns also contribute to uncertainties in orthology assignment, alignment, and data analysis. The targets of the method described here are well characterized, protein-coding orthologs that are easy to align and predisposed to evolutionary analysis with little bioinformatic preprocessing.

Gene capture technology allows researchers to target, isolate, and sequence hundreds or thousands of genes of interest from genomic libraries. As powerful as it is, the technique is currently restricted in scope to comparisons within species, among species with limited divergence or by targeting highly conserved fragments across otherwise divergent taxa. The modifications presented here extend the power of gene capture to comparisons among highly divergent sequences and taxa. This has obvious implications for molecular systematics and evolutionary genetics. Perhaps more importantly, our approach holds promise for comparative biochemistry, physiology and medicine as it will expand the range of evolutionary comparisons that can be efficiently explored for pre-specified genes associated with particular biochemical pathways, physiological adaptations, and disease conditions.

Acknowledgments

This research was funded by the National Science Foundation (DEB award -1132229) to GJPN. CL is also supported by the Leading Academic Discipline Project of the Shanghai Municipal Education Commission, project number: S30701. We thank Johanna L. A. Paijmans for technical assistance in the lab.

Competing interests

The authors declare no competing interests.

References

1. **Sulem, P., D.F. Gudbjartsson, S.N. Stacey, A. Helgason, T. Rafnar, M. Jakobsdottir, S. Steinberg, S.A. Gudjonsson, et al.** 2008. Two newly identified genetic determinants of pigmentation in Europeans. Nat. Genet. *40*:835-837.

2. **Ng, S.B., E.H. Turner, P.D. Robertson, S.D. Flygare, A.W. Bigham, C. Lee, T. Shaffer, M. Wong, et al.** 2009. Targeted capture and massively parallel sequencing of 12 human exomes. Nature *461*:272-276.
3. **Jones, F.C., M.G. Grabherr, Y.F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, et al.** 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature *484*:55-61.
4. **Vasta, V., S.B. Ng, E.H. Turner, J. Shendure, and S.H. Hahn.** 2009. Next generation sequence analysis for mitochondrial disorders. Genome Med. *1*:100.
5. **Hofreiter, M. and T. Schoneberg.** 2010. The genetic and evolutionary basis of colour variation in vertebrates. Cell. Mol. Life Sci. *67*:2591-2603.
6. **Kalujnaia, S., I.S. McWilliam, V.A. Zaguinaiko, A.L. Feilen, J. Nicholson, N. Hazon, C.P. Cutler, R.J. Balment, et al.** 2007. Salinity adaptation and gene profiling analysis in the European eel (Anguilla anguilla) using microarray technology. Gen. Comp. Endocrinol. *152*:274-280.
7. **Hickey, A.J., G.M. Renshaw, B. Speers-Roesch, J.G. Richards, Y. Wang, A.P. Farrell, and C.J. Brauner.** 2012. A radical approach to beating hypoxia: depressed free radical release from heart fibres of the hypoxia-tolerant epaulette shark (Hemiscyllum ocellatum). J. Comp. Physiol. B *182*:91-100.
8. **Hodges, E., Z. Xuan, V. Balija, M. Kramer, M.N. Molla, S.W. Smith, C.M. Middle, M.J. Rodesch, et al.** 2007. Genome-wide in situ exon capture for selective resequencing. Nat. Genet. *39*:1522-1527.
9. **Gnirke, A., A. Melnikov, J. Maguire, P. Rogov, E.M. LeProust, W. Brockman, T. Fennell, G. Giannoukos, et al.** 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat. Biotechnol. *27*:182-189.
10. **Mason, V.C., G. Li, K.M. Helgen, and W.J. Murphy.** 2011. Efficient cross-species capture hybridization and next-generation sequencing of mitochondrial genomes from noninvasively sampled museum specimens. Genome Res. *21*:1695-1704.
11. **Cosart, T., A. Beja-Pereira, S. Chen, S.B. Ng, J. Shendure, and G. Luikart.** 2011. Exome-wide DNA capture and next generation sequencing in domestic and wild species. BMC Genomics *12*:347.
12. **Bi, K., D. Vanderpool, S. Singhal, T. Linderoth, C. Moritz, and J.M. Good.** 2012. Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. BMC Genomics *13*:403.
13. **Hancock-Hanser, B.L., A. Frey, M.S. Leslie, P.H. Dutton, F.I. Archer, and P.A. Morin.** 2013. Targeted multiplex next-generation sequencing: advances in techniques of mitochondrial and nuclear DNA sequencing for population genomics. Mol Ecol Resour *13*:254-268.
14. **McCormack, J.E., B.C. Faircloth, N.G. Crawford, P.A. Gowaty, R.T. Brumfield, and T.C. Glenn.** 2012. Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. Genome Res. *22*:746-754.
15. **Lemmon, A.R., S.A. Emme, and E.M. Lemmon.** 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst. Biol. *61*:727-744.
16. **Southern, E.M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. *98*:503-517.
17. **Altschul, S.F., W. Gish, W. Miller, E. Myers, and D. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. *215*:403-410.
18. **Li, C., J.J. Riethoven, and G.J.P. Naylor.** 2012. EvolMarkers: a database for mining exon and intron markers for evolution, ecology and conservation studies. Mol Ecol Resour. *12*:967-971.
19. **Inoue, J.G., M. Miya, K. Lam, B.H. Tay, J.A. Danks, J. Bell, T.I. Walker, and B. Venkatesh.** 2010. Evolutionary Origin and Phylogeny of the Modern Holocephalans (Chondrichthyes: Chimaeriformes): A Mitogenomic Perspective. Mol Biol Evol.
20. **Martin, M.** 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*:10-12.
21. **Simpson, J.T., K. Wong, S.D. Jackman, J.E. Schein, S.J. Jones, and I. Birol.** 2009. ABySS: a parallel assembler for short read sequence data. Genome Res. *19*:1117-1123.
22. **Ebersberger, I., S. Strauss, and A. von Haeseler.** 2009. HaMStR: profile hidden markov model based search for orthologs in ESTs. BMC Evol. Biol. *9*:157.
23. **Avise, J.C., J.C. Patton, and C.F. Aquadro.** 1980. Evolutionary genetics of birds. Comparative molecular evolution in New World warblers and rodents. J. Hered. *71*:303-310.
24. **Straus, N.A.** 1971. Comparative DNA renaturation kinetics in amphibians. Proc. Natl. Acad. Sci. USA *68*:799-802.
25. **Fu, Y., N.M. Springer, D.J. Gerhardt, K. Ying, C.T. Yeh, W. Wu, R. Swanson-Wagner, M. D'Ascenzo, et al.** 2010. Repeat subtraction-mediated sequence capture from a complex genome. Plant J. *62*:898-909.
26. **Lane, S., J. Evermann, F. Loge, and D.R. Call.** 2004. Amplicon secondary structure prevents target hybridization to oligonucleotide microarrays. Biosens. Bioelectron. *20*:728-735.
27. **Crawford, N.G., B.C. Faircloth, J.E. McCormack, R.T. Brumfield, K. Winker, and T.C. Glenn.** 2012. More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. Biol. Lett. *8*:783-786.
28. **Faircloth, B.C., J.E. McCormack, N.G. Crawford, M.G. Harvey, R.T. Brumfield, and T.C. Glenn.** 2012. Ultraconserved Elements Anchor Thousands of Genetic Markers Spanning Multiple Evolutionary Timescales. Systematic biology.
29. **Hedges, S.B. and S. Kumar.** 2009. The TimeTree of Life. Oxford University Press, New York.

Received 26 February 2013; accepted 25 April 2013.

Address correspondence to Gavin J P Naylor, Hollings Marine Laboratory, College of Charleston, Charleston, SC, USA. E-mail: naylorg@cofc.edu.