KOREAN JOURNAL OF ICHTHYOLOGY, Vol. 33, No. 4, 205-216, December 2021
ISSN: 1225-8598 (Print), 2288-3371 (Online). DOI: https://doi.org/10.35399/ISK.33.4.1

Received: October 25, 2021
Revised: November 29, 2021
Accepted: November 29, 2021

Exon Capture - Principle and Applications to Phylogenomics and Population Genomics of Fishes

By Chenhong Li[1,2,*]

[1]Shanghai Universities Key Laboratory of Marine Animal Taxonomy and Evolution, Shanghai Ocean University, Shanghai 201306, China
[2]Shanghai Collaborative Innovation for Aquatic Animal Genetics and Breeding, Shanghai Ocean University, Shanghai 201306, China

ABSTRACT **Phylogenetic reconstruction based on one locus or a few loci can be misleading due to gene-tree/species-tree discordance. Species delimitation and intraspecific studies also often suffered from low resolution because of insufficient statistic power when few loci were used. Exon capture method is one of the most efficient way to collect genome-scale data, which can significantly augment studies that aimed to investigate patterns and histories of organisms at both intraspecific and high level. Here, I showed the advancement of shifting from single-gene method to genomic approach and the benefit of applying exon capture method comparing to alternative genomic techniques. Then, I explained the principle of exon capture method as well as providing detailed recommendations for applying this method. Finally, I demonstrated exon capture method using two applications and discussed future perspectives of this technology.**

Key words: Target enrichment, exon capture, species delimitation, population structure, population dynamics, adaptation, molecular systematics, phylogenetics

INTRODUCTION

Because there are more molecular data available than morphological characters, and homology of molecular data usually is easier to be identified than the morphological characters, molecular data have become the predominant evidence to reconstruct tree of life. Molecular data also have been wildly used as the indispensable source for studying population structure and dynamics. Traditional molecular markers include mitochondrial genes or a few nuclear loci. Gene genealogy estimated from a single gene does not necessarily consistent with evolutionary history of species. Processes, such as gene introgression and incomplete lineage sorting can mislead species-level phylogenetic reconstruction based on single gene. For example, in our study of the Chinese perch (*Siniperca* spp.) using Cytochrome c oxidase I (COI) gene, we found that different species were mixed together on the tree (Liu *et al*., 2017). One individual of *S. obscura* was more closely related to *S. kneri* than to other individuals of *S. obscura*, and individuals of the *S. kneri* and *S. chuatsi* were all mixed together (Fig. 1), but the problems were resolved when more independent nuclear loci were used (Liu *et al*., 2017).

High-level phylogenies based on single locus also can be misleading. In my dissertation work (Li, 2007), I showed that maximum likelihood phylogenies based on single-copy nuclear genes were not consistent with each other, that is, no two trees had the same topology (Fig. 2). Paralogy (Maddison, 1997), incomplete lineage sorting (Funk and Omland, 2003; Maddison and Knowles, 2006), horizontal gene transfer (Kurland *et al*., 2003) and stochastic errors all can led to the inconsistent results. Using genome-scale data can help to sort out nonphylogenetic noise and recover the true phylogenetic signals. With a large number of characters, the stochastic errors associated with the estimations decreases (Delsuc *et al*., 2005), and many independent nuclear genes can reduce some systematic errors (Collins *et al*., 2005; Maddison and Knowles, 2006).

Species delimitation methods, such as DNA barcoding

저자 직위: Chenhong Li (교수)
*Corresponding author: Chenhong Li Fax: 86-21-61900263,
E-mail: chli@shou.edu.cn

also are based on single locus, most commonly COI gene. We have showed that single locus sometimes cannot distinguish recently diverged species or sister species if there was mild gene flow between them (Liu *et al*., 2017). The genetic distance between species was not greater than the genetic distance between individuals within species when one locus or a few loci were used. When more independent loci were added, the species can be correctly distinguished and a clear gap between the within-species distance and the between-species distance revealed (Fig. 3).

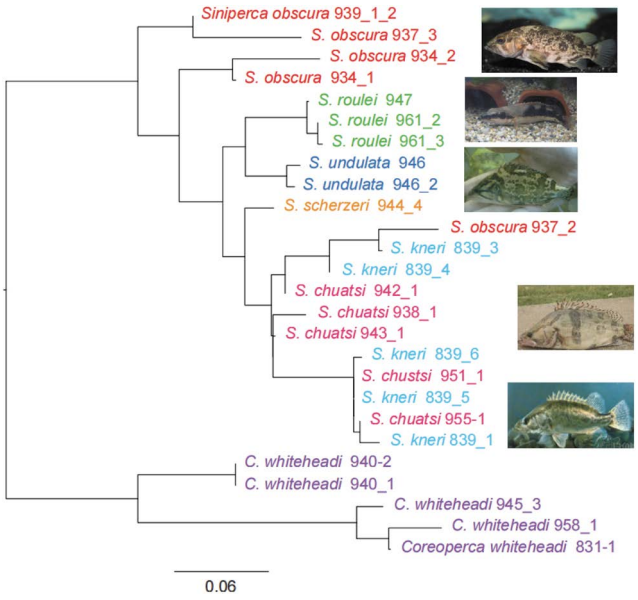

Fig. 1. Phylogeny of the Chinese perches (genus: *Siniperca*) based on Cytochrome c oxidase I gene (modified from Liu *et al*., 2017).

As showed above, using a single gene or a few loci to study evolutionary history and population genetics of fishes often lead to erroneous results at both species and higher level. Using genome-scale data can improve estimation of population parameters (Luikart *et al*., 2003) and resolve inconsistency in phylogenetic reconstruction due to gene-tree/species-tree discordance (Li, 2007).

EXON CAPTURE AND OTHER GENOMIC APPROACHES FOR POPULATION GENETICS AND PHYLOGENOMICS

There many ways to acquire genome-scale data for population genetics and phylogenomic studies. The straight-

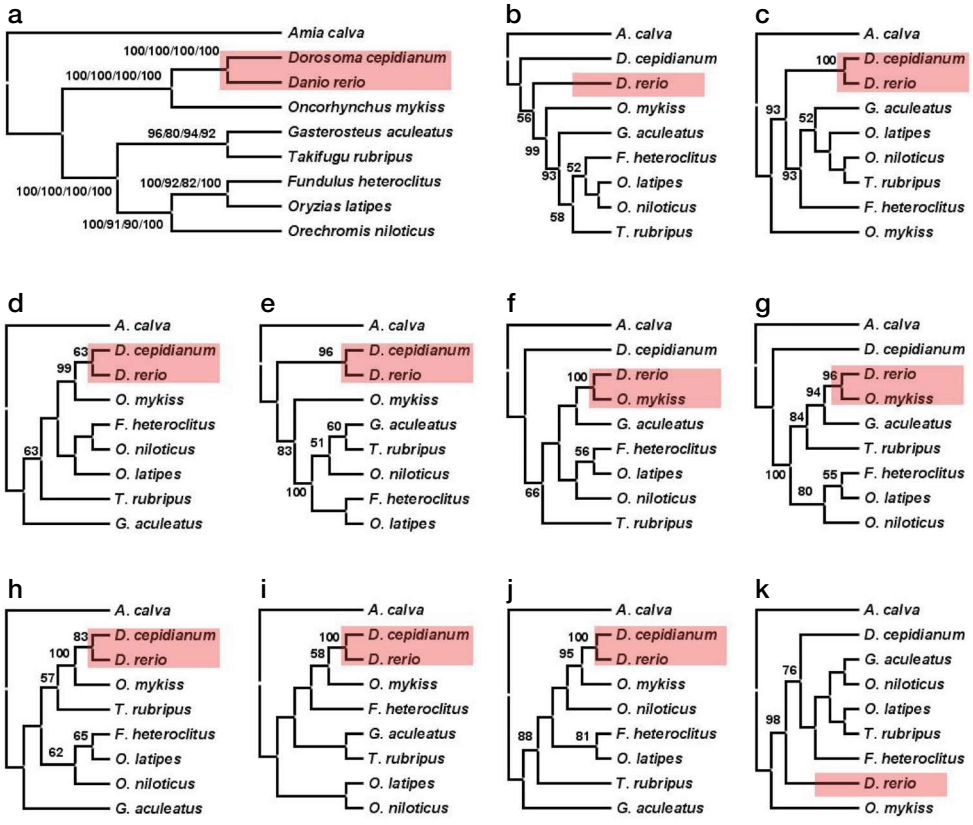

Fig. 2. Maximum likelihood phylogenies of teleosts based on nuclear genes: zic1, myh6, RYR3, Ptc, tbr1, ENC1, Gylt, SH3PX3, plagl2, and sreb2.

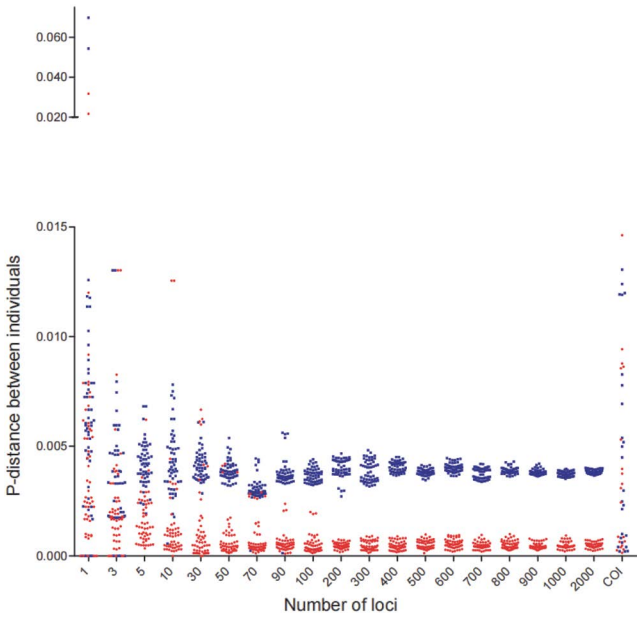

Fig. 3. Within-species genetic distance (red) and between-species genetic distance (blue) of *Siniperca chuatsi* and *S. kneri*. The distance was calculated using 1 to 2000 nuclear loci or using the Cytochrome c oxidase I (COI) gene (modified from Liu *et al*., 2017).

forward way is to sequence the whole genome, such as reconstructing phylogeny of birds (Jarvis *et al*., 2014) and ruminants (Chen *et al*., 2019) using whole genome data. Nonetheless, it is still costly to sequence whole genome for population studies or phylogenetics and it is not necessary most of the time.

Reduced-representation sequencing is a technique used to extract parts of the genome, often orthologous in different samples and then make those into sequencing libraries for the next generation sequencing. It has been wildly used for phylogenetics and population studies in recent years (Emerson *et al*., 2010; Faircloth *et al*., 2012; Lemmon *et al*., 2012; Li *et al*., 2013). Two strategies are commonly found in reduced-representation sequencing. First is related to restriction site, such as Restriction-site Associated DNA (RAD) Sequencing (Baird *et al*., 2008) and double digest RADseq (ddRAD) markers (Peterson *et al*., 2012). These methods could be used to generate a large amount of data from anonymous loci that are particularly useful in studying population genomics or evolution history under species level (Davey and Blaxter, 2010), but homology of the anonymous loci becomes uncertain, when divergent species were compared. Furthermore, for each species, a new set of loci need to be developed, which makes comparison between studies difficult. The other strategy of reduced-representation sequencing is gene capture, also known as target enrichment, which often result in less missing data than the restriction site-related methods do (Collins and Hrbek, 2015). Target loci can be applied across highly divergent taxonomic groups (Faircloth *et al*., 2012; Lemmon *et al*., 2012; Li *et al*., 2013). Gene-capture methods can be used to enrich highly anonymous conserved regions and utilize variable flanking regions for data analyses, such as the method of Ultraconserved Element (UCE) captures (Faircloth *et al*., 2012) and Anchored Hybrid Enrichment (AHE) (Lemmon *et al*., 2012). Nevertheless, gene capture also can be used to target exons directly (Bi *et al*., 2012; Li *et al*., 2012; Hedtke *et al*., 2013).

Exons have been more commonly used for phylogenetics than anonymous noncoding regions, and evolution propriety of protein-coding sequences has been well studied. Furthermore, the flanking region of exons make these markers can also be used for population genetic studies. The papers listed in Table 1 were retrieved from PubMed (https://pubmed.ncbi.nlm.nih.gov/, accessed on Oct 15, 2021) using the keywords: gene + capture + fish for publications from 2011 to 2021. There are 29 papers published on fish population genetics and phylogenomics using exon capture methods and two papers on studying functional genes. These studies covered both osteichthyans and chondrichthyans. Eighteen papers were focused on high-level phylogenetics, and 11 papers on population genetics or phylogeographics. Application of exon capture methods on population-level studies is yet fully exploited.

PRINCIPLE OF EXON CAPTURE METHOD AND RECOMMENDATIONS

Exon capture is based on hybridizing RNA/DNA baits (probes) to DNA libraries of targeted species and enriching sequences similar to the baits for subsequent high-throughput sequencing. Exon capture method involves three major aspects: target selection and baits design, library preparation, and gene capture and amplification.

Exon capture markers have been developed for osteichthyans as well as chondrichthyans (Li *et al*., 2013; Ilves and López-Fernández, 2014; Nielsen *et al*., 2017; Jiang *et al*., 2019; Hughes *et al*., 2021). Ilves and López-Fernández (2014) developed 923 exon markers for cichlids based on genome sequence of *Oreochromis niloticus*. They compared single-copy exons found in *O. niloticus* with genomic data of four additional African cichlid species and chose the single-copy exons across all five cichlid species as their targets. Those markers worked well and were applied in

Table 1. A sample of recent studies (2011−2021) on phylogenomics and population genetics of fishes using exon capture methods. Articles were retrieved from PubMed (https://pubmed.ncbi.nlm.nih.gov/, accessed on Oct 15, 2021) using the keywords: gene + capture + fish

Authors and year	Title	Taxonomic level/taxa	Research directions
Corrigan *et al*., 2017	Historical introgression drives pervasive mitochondrial admixture between two species of pelagic sharks	Species/*Carcharhinus*	Population species delimitation
Liu *et al*., 2017	Multilocus DNA barcoding - species identification with multilocus data	Species/*Siniperca*	Population species delimitation
Rincon-Sandoval *et al*., 2019	Comparative phylogeography of trans-Andean freshwater fishes based on genome-wide nuclear and mitochondrial markers	Species/ostariophysans	Population phylogeography
Therkildsen and Palumbi, 2017	Practical low-coverage genome-wide sequencing of hundreds of individually barcoded samples for population and evolutionary genomics in non-model species	Species/the Atlantic silverside	Population (in silico exon capture)
Ai *et al*., 2021	Genetic and morphological differences between yellowtail kingfish (*Seriola lalandi*) from the Bohai Sea, China and the Southern Ocean, Australia	Species/the yellowtail kingfish	Population genetics
Cheng *et al*., 2019	Multiple freshwater invasions of the tapertail anchovy (Clupeiformes: Engraulidae) of the Yangtze River	Species/the tapertail anchovy	Population genetics
Nielsen *et al*., 2017	Extracting DNA from 'jaws': high yield and quality from archived tiger shark (*Galeocerdo cuvier*) skeletal material	Species/the tiger shark	Population genetics
Sarker *et al*., 2020	Genetic diversity of *Hilsa kelee* collected from the Bay of Bengal and the Arabian Sea	Species/*Hilsa kelee*	Population genetics
Sarker *et al*.,2021	Cross-species gene enrichment revealed a single population of Hilsa shad (*Tenualosa ilisha*) with low genetic variation in Bangladesh waters	Species/Hilsa shad	Population genetics
Li *et al*., 2015	DNA capture reveals transoceanic gene flow in endangered river sharks	Species/*Glyphis*	Population dynamics
Maisano Delser *et al*., 2016	Population genomics of *C. melanopterus* using target gene capture data: demographic inferences and conservation perspectives	Species/the blacktip reef shark	Population dynamics
Arcila *et al*., 2021	Testing the utility of alternative metrics of branch support to address the ancient evolutionary radiation of tunas, stromateoids, and allies (Teleostei: Pelagiaria)	Order/pelagiarian fishes	Phylogenomics
Atta *et al*., 2021	Exon-capture data and locus screening provide new insights into the phylogeny of flatfishes (Pleuronectoidei)	Superorder/flatfishes	Phylogenomics
Betancur *et al*., 2019	Phylogenomic incongruence, hypothesis testing, and taxonomic sampling: The monophyly of characiform fishes	Family/Characiformes	Phylogenomics
Campbell *et al*., 2020	Addressing incomplete lineage sorting and paralogy in the inference of uncertain salmonid phylogenetic relationships	Family/Salmonidae	Phylogenomics

Table 1. Continued

Authors and year	Title	Taxonomic level/taxa	Research directions
Hughes *et al*., 2021	Exon probe sets and bioinformatics pipelines for all levels of fish phylogenomics	All level/ray-finned fishes	Phylogenomics
Ilves and López-Fernández, 2014	A targeted next-generation sequencing toolkit for exon-based cichlid phylogenomics	Family/Cichilidae	Phylogenomics
Ilves *et al*., 2018	Exon-based phylogenomics strengthens the phylogeny of Neotropical cichlids and identifies remaining conflicting clades (Cichliformes: Cichlidae: Cichlinae)	Family/Cichilidae	Phylogenomics
Jiang *et al*., 2019	Gene markers for exon capture and phylogenomics in ray-finned fishes	All level/ray-finned fishes	Phylogenomics
Kolmann *et al*., 2021	Phylogenomics of piranhas and pacus (Serrasalmidae) uncovers how convergent diets obfuscate traditional morphological taxonomy	Family/Serrasalmidae	Phylogenomics
Kuang *et al*., 2018	Phylogenomic analysis on the exceptionally diverse fish clade Gobioidei (Actinopterygii: Gobiiformes) and data-filtering based on molecular clocklikeness	Order/Gobiiformes	Phylogenomics
Li *et al*., 2013	Capturing protein-coding genes across highly divergent species	Phylum/Vertebrata	Phylogenomics
Li *et al*., 2018	Molecular systematics and phylogenetic analysis of the Asian endemic freshwater sleepers (Gobiiformes: Odontobutidae)	Family/Odontobutidae	Phylogenomics
Roa-Varón *et al*., 2021	Confronting sources of systematic error to resolve historically contentious relationships: a case study using gadiform fishes (Teleostei, Paracanthopterygii, Gadiformes)	Order/Gadiformes	Phylogenomics
Song *et al*., 2017	Species delimitation and phylogenetic reconstruction of the sinipercids (Perciformes: Sinipercidae) based on target enrichment of thousands of nuclear coding sequences	Family/Sinipercidae	Phylogenomics
Straube *et al*., 2015	Molecular phylogeny of Squaliformes and first occurrence of bioluminescence in sharks	Family/Squaliformes	Phylogenomics
Straube *et al*., 2018	A phylogenomic approach to reconstruct interrelationships of main clupeocephalan lineages with a critical discussion of morphological apomorphies	Superorder/Clupeocephalan	Phylogenomics
White *et al*., 2018	Phylogeny of the manta and devilrays (Chondrichthyes: Mobulidae), with an updated taxonomic arrangement for the family	Family/Chondrichthyes: Mobulidae	Phylogenomics
Yin *et al*., 2019	Molecular systematics of *Pampus* (Perciformes: Stromateidae) based on thousands of nuclear loci using target-gene enrichment	Genus/*Pampus*	Phylogenomics
Hebert *et al*., 2013	Targeted sequence capture and resequencing implies a predominant role of regulatory regions in the divergence of a sympatric lake whitefish species pair (*Coregonus clupeaformis*)	Species/lake white fish	Functional genes
Lappin *et al*., 2016	Targeted sequencing for high-resolution evolutionary analyses following genome duplication in salmonid fish: Proof of concept for key components of the insulin-like growth factor axis	Family/Salmonidae	Functional genes

study of Neotropical cichlids (Ilves *et al*., 2018). However, their exon markers were selected to target only cichlids, not intended to be used for other fish species. Moreover, they chose target exons that were relatively long (750−2000 bp), so limited the number of potential exon markers, because most exons of vertebrates are around 200 bp (Li *et al*., 2013). Similarly, taxon-specific exon markers were used for studying archived tiger shark (*Galeocerdo cuvier*). Nielsen *et al*. (2017) designed baits on 44,794 transcriptome sequences of the lesser spotted catshark (*Scyliorhinus canicula*) and used those to test gene capture on archived tiger shark samples.

Taxon-specific exon markers not only limit the potential applicable taxonomic range of these markers, but also make the results not comparable or integrable from different fish lineages. Exon-capture targets for all vertebrates were explored (Li *et al*., 2013). Genomes of six vertebrate: human (*Homo sapiens*), chicken (*Gallus gallus*), western clawed toad (*Xenopus tropicalis*), green anole (*Anolis carolinensis*), zebrafish (*Danio rerio*), and elephant shark (*Callorhinchus milii*) were compared to identify putatively orthologous genes within each respective genome. A total of 1,449 exon loci were selected and tested in different vertebrate lineages with various success rate (Li *et al*., 2013). Nonetheless, few studies applied those "vertebrate markers" on fish studies, probably due to the small number of applicable loci and that those are not fish-specific markers. Subsequently, Song *et al*. (2017) developed 17,817 single-copy nuclear coding sequence (CDS) markers for percomorph fishes and applied those in the sinipercids with great success. Hughes *et al*. (2018) selected 1,721 exon markers $>$200 bp from the 17,817 markers and retrieved their sequences from hundreds of transcriptomic and genomic datasets in silico. Recently, exon capture markers for all ray-finned fishes were optimized and tested (Jiang *et al*., 2019; Hughes *et al*., 2021). Exon capture markers specific for all chondrichthyans are yet to be developed.

Exon markers can be identified by comparing genomes of model species (Fig. 4). Many considerations were taken into baits design, such as uniqueness and conservativeness of markers, length and complexity of markers, and genetic distance between baits and target sequences (Bi *et al*., 2012; Faircloth *et al*., 2012; Lemmon *et al*., 2012; Li *et al*., 2013; Hugall *et al*., 2016; Campana, 2018). I have four recommendations for selecting exon capture markers: (1) single-copy genes are preferred, because they are less susceptible to paralogy problem; (2) relax the length requirement for exon markers, because most exons of vertebrates are short; (3) final set of in silico identified markers should be tested empirically. Once the target exons have been selected, their

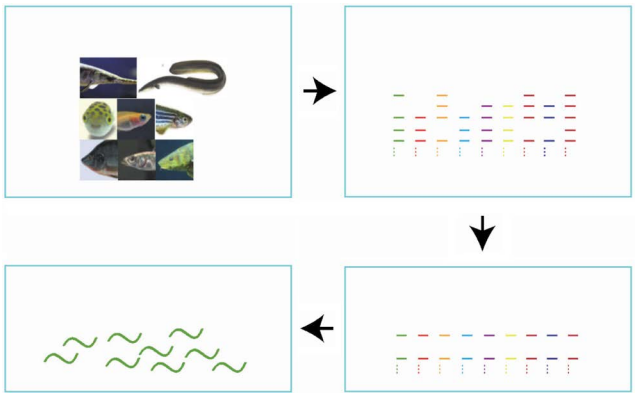

Fig. 4. Pipeline for developing single-copy conserved exon markers for ray-finned fishes. Single-copy exons were identified in each genome; conserved single-copy exons from all compared species were selected for designing RNA baits.

sequence information can be send to company for refined bait designing and synthesizing, such as ArborBiosciences (Ann Arbor, MI, USA); (4) use the sequences of a close relative to your species of interest for bait design if available.

The second aspect of exon capture is library preparation, which is the process of fragmenting DNA of target species and adding adapters to both end of the fragments. There are many ways for library preparation. The most commonly used protocol involves ultrasonic shearing DNA, blunt end, ligation and fill-in steps (Meyer and Kircher 2010; Huang *et al*., 2021). Other protocols, such as tagmentation library prep (https://www.illumina.com/techniques/sequencing/ngs-library-prep/tagmentation.html) involves less hands-on time, but those commercial kits are very expensive. Using the common protocol and recipe for homemade reagents can significantly lower the cost of exon capture (Huang *et al*., 2021). Usually 30−300 ng starting materials are enough for library prep, and extra DNA would not significantly increase the product. If there are only trace amount of DNA available, pre-amplification genomic DNA using MALBAC method can be applied (Zong *et al*., 2012). Size of inserted DNA usually is around 500 bp. If fragments larger than 1000 bp were detectable, size selection should be carried out to remove those long fragments to improve sequencing results. Eighteen cycles of the last amplification step is recommended, because PCR duplication is less of concern than inadequate material before gene capture steps.

The gene capture steps should follow manual of commercial kits, such as myBaits (https://arborbiosci.com/wp-content/uploads/2021/03/myBaits_v5.01_Manual.pdf). Since the number of targeting loci usually is much lower than

regular whole-exome capture, one tenth of the bait recommended volume could be used. Because fish exon capture usually is applied across species, a touch-down temperature scheme should be used and the washing temperature must be lowered accordingly (Li *et al*., 2013). During the hybridization and washing step, we should keep in mind that the purpose of this procedure is to separate the target DNA fragments from non-specific ones, so precise temperature control should be managed and discarding unwanted should be complete.

APPLICATIONS APPLYING EXON CAPTURE METHOD AND FUTURE PERSPECTIVE

With the development of exon-capture markers and lowering cost of library preparation and sequencing, more and more phylogenomic and population genomic studies involved exon capture approaches (Maisano Delser *et al*., 2016; Corrigan *et al*., 2017; Liu *et al*., 2017; Song *et al*., 2017; Kuang *et al*., 2018; Li *et al*., 2018; Straube *et al*., 2018; White *et al*., 2018; Betancur *et al*., 2019; Cheng *et al*., 2019; Rincon-Sandoval *et al*., 2019; Yin *et al*., 2019; Sarker *et al*., 2020; Ai *et al*., 2021; Arcila *et al*., 2021; Atta *et al*., 2021; Kolmann *et al*., 2021; Roa-Varón *et al*., 2021; Sarker *et al*., 2021). Here I illustrate what we can do with exon capture methods using two examples from my lab.

The Odontobutidae is a group of sleeper fishes with six genera and 15−22 species. The composition of the Odontobutidae and the interrelationship of the odontobutids were unresolved. We collected sequence data of 4,434 single-copy exons from 41 specimens of odontobutids and reconstructed a robust phylogeny of the Odontobutidae (Li *et al*., 2018). After removing PCR duplicates, in average, 57% were unique reads. After discarding loci that cannot be aligned well, 4,397 were left for further analyses. Phylogenetic tree had 100% bootstrap support for all nodes, and species tree generated the same result (Fig. 5). One hundred most clock-like loci were used to calibrate a time tree. The most recent common ancestor of odontobutids was estimated at 30.8 Ma (20.7−41.9 Ma, 95% HPDs). Additionally, DEC analysis implemented in RASP showed that ancestor of the odontobutids was distributed in southern China and Indo-China Peninsula (Li *et al*., 2018).

In another study, we tested hypotheses about freshwater invasion of *Coilia nasus* and identified loci adapted to non-migratory freshwater habitat (Cheng *et al*., 2019). We captured 4,434 exon for the *Coilia* species complex. Read assembly and data filtering produced 1,813 loci for each sample on average, and 2,869 clean target were kept for subsequent analyses. We found that both *C. nasus* and *C. brachygnathus* were valid species using Bayes factor species delimitation (BFD*). Two independent freshwater invasion events with subsequent gene flow between adjacent populations were supported by fastsimcoal2 analyses, with the first event occurring around 4.07 Ma and the second happened around 3.2 Ka (Fig. 6). F-DIST analyses singled out 120 outliers by comparing migratory *C. nasus* and *C. brachygnathus*, and 21 outlier between migratory and landlocked *C. nasus*. Nine of those loci were shared between the two comparisons, suggesting that those might play a conserved and important role in adaptation of *C. nasus* to freshwater habitat (Cheng *et al*., 2019). Furthermore, population structure and migration between populations were estimated using the exon-capture data.

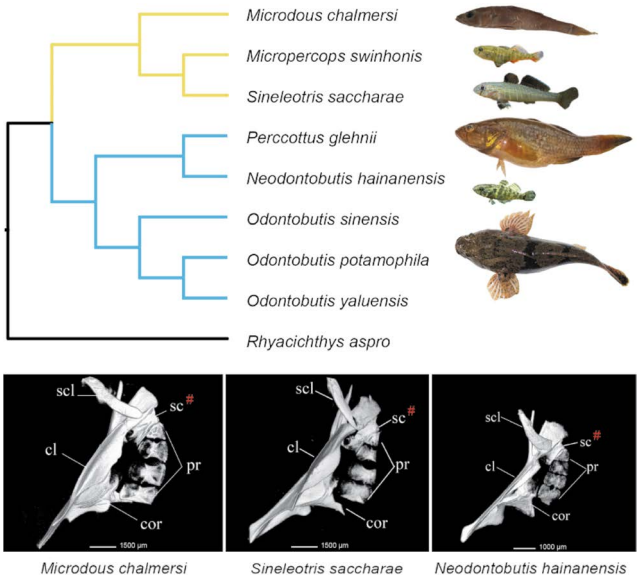

Fig. 5. Phylogenomic study on the Odontobutidae, modified from Li *et al*. (2018).

The two examples illustrated that exon-capture data can be used to trace evolutionary history at both specie level and higher level. It can also be used to identify adaptive loci given good comparative setting. Given the decreasing cost for sequencing, one might argue that sequencing the whole genome is easier than exon capture method. Nonetheless, the cost is still higher for whole genome sequencing than exon capture, particular for many fish species that have large genome size. Furthermore, for archive samples or samples with contamination from environment and low proportion of endogenous DNA, exon capture is the best way to enrich the target loci. Finally, gene capture can be used to enrich both mitochondrial and nuclear DNA fragments in environmental DNA research (Jensen *et al*., 2021).

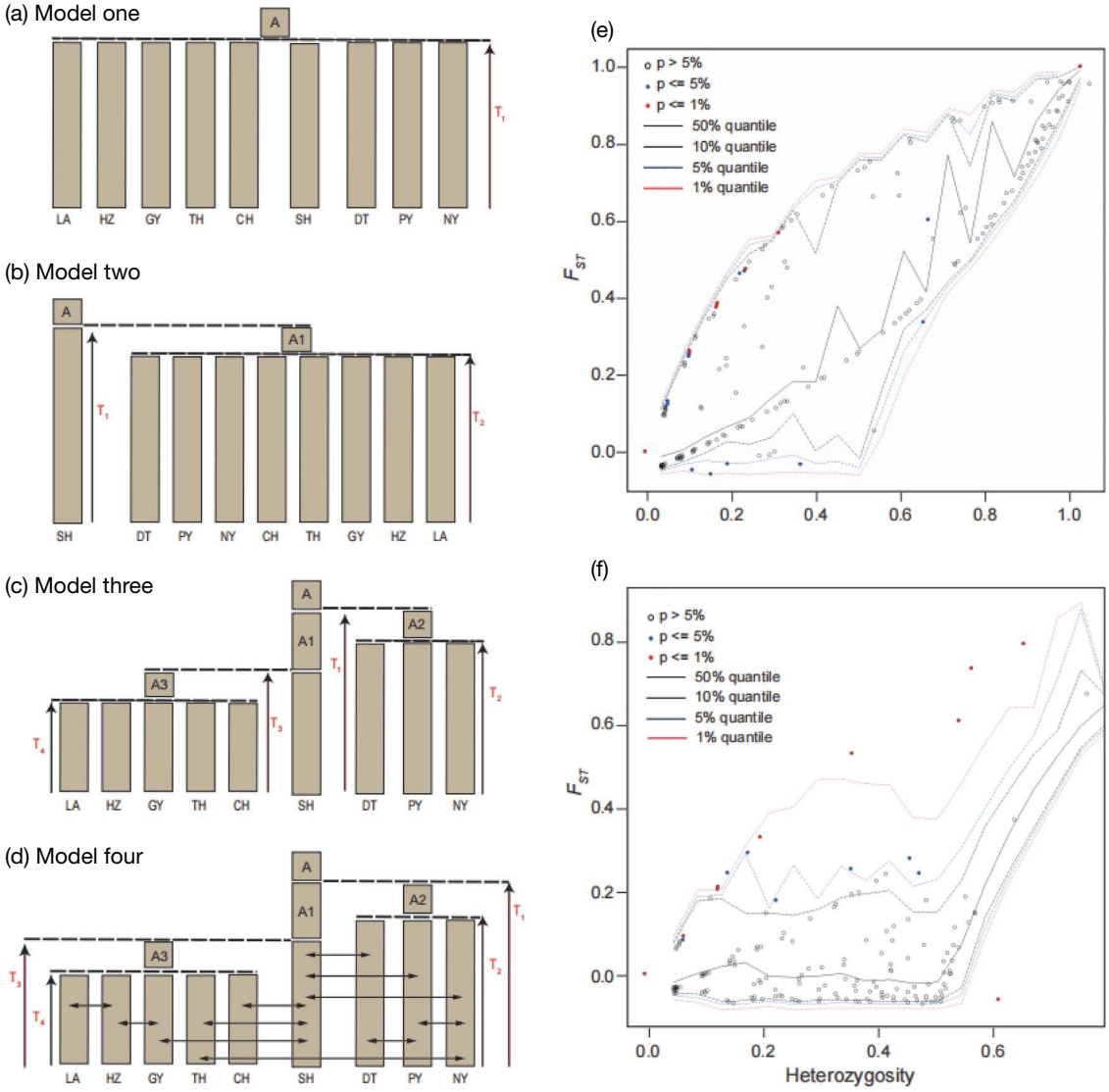

Fig. 6. Population genomics of *Coilia nasus*, modified from Cheng *et al*. (2019). Left, model test using fastsimcoal2 supported twice freshwater invasion of *C. nasus* followed by gene flow between adjacent populations; right, F-dist test revealed 21 outlier loci between migratory population and landlocked *C. nasus* and 120 outlier loci between the migratory population of *C. nasus* and *C. brachygnathus*.

ACKNOWLEDGEMENTS

CL was supported by the Science and Technology Commission of Shanghai Municipality (19050501900, 19410 740500).

REFERENCES

Ai, Q., L. Sang, H. Tan, X. Huang, B. Bao and C. Li. 2021. Genetic and morphological differences between yellowtail kingfish (*Seriola lalandi*) from the Bohai Sea, China and the Southern Ocean, Australia. Aquac. Fish., 6: 260-266. https://doi.org/10.1016/j.aaf.2020.03.004.

Arcila, D., L.C. Hughes, B. Meléndez-Vazquez, C.C. Baldwin, W.T. White, K.E. Carpenter, J.T. Williams, M.D. Santos, J.J. Pogonoski, M. Miya, G. Ortí and R.R. Betancur. 2021. Testing the Utility of alternative metrics of branch support to address the ancient evolutionary radiation of Tunas, Stromateoids, and Allies (Teleostei: Pelagiaria). Syst. Biol., 70: 1123-1144. https://doi.org/10.1093/sysbio/syab018.

Atta, C.J., H. Yuan, C. Li, D. Arcila, R.R. Betancur, L.C. Hughes, G. Ortí and L. Tornabene. 2021. Exon-capture data and locus screening provide new insights into the phylogeny of flatfishes (Pleuronectoidei). Mol. Phylogenet. Evol., 166: 107315. https://doi.org/10.1016/j.ympev.2021.107315.

Baird, N.A., P.D. Etter, T.S. Atwood, M.C. Currey, A.L. Shiver, Z.A. Lewis, E.U. Selker, W.A. Cresko and E.A. Johnson. 2008.

Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE, 3: e3376. https://doi.org/10.1371/journal.pone.0003376.

Betancur, R.R., D. Arcila, R.P. Vari, L.C. Hughes, C. Oliveira, M.H. Sabaj and G. Ortí. 2019. Phylogenomic incongruence, hypothesis testing, and taxonomic sampling: The monophyly of characiform fishes. Evolution, 73: 329-345. https://doi.org/10.1111/evo.13666.

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 Genom., 13: 403. https://doi.org/10.1186/1471-2164-13-403.

Campana, M.G. 2018. BaitsTools: Software for hybridization capture bait design. Mol. Ecol. Resour., 8: 356-361. https://doi.org/10.1111/1755-0998.12721.

Campbell, M.A., T.J. Buser, M.E. Alfaro and J.A. López. 2020. Addressing incomplete lineage sorting and paralogy in the inference of uncertain salmonid phylogenetic relationships. PeerJ, 8: e9389. https://doi.org/10.7717/peerj.9389.

Chen, L., Q. Qiu, Y. Jiang, K. Wang, Z. Lin, Z. Li, F. Bibi, Y. Yang, J. Wang, W. Nie, W. Su, G. Liu, Q. Li, W. Fu, X. Pan, C. Liu, J. Yang, C. Zhang, Y. Yin, Y. Wang, Y. Zhao, C. Zhang, Z. Wang, Y. Qin, W. Liu, B. Wang, Y. Ren, R. Zhang, Y. Zeng, R.R. da Fonseca, B. Wei, R. Li, W. Wan, R. Zhao, W. Zhu, Y. Wang, S. Duan, Y. Gao, Y.E. Zhang, C. Chen, C. Hvilsom, C.W. Epps, L.G. Chemnick, Y. Dong, S. Mirarab, H.R. Siegismund, O.A. Ryder, M.T.P. Gilbert, H.A. Lewin, G. Zhang, R. Heller and W. Wang. 2019. Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. Science, 364: eaav6202. https://doi.org/10.1126/science.aav6202.

Cheng, F., Q. Wang, D.P. Maisano and C. Li. 2019. Multiple freshwater invasions of the tapertail anchovy (Clupeiformes: Engraulidae) of the Yangtze River. Ecol. Evol., 9, 12202-12215. https://doi.org/10.1002/ece3.5708.

Collins, R.A. and T. Hrbek. 2015. An in silico comparison of reduced-representation and sequence-capture protocols for phylogenomics. BioRxiv, 032565. https://doi.org/10.1101/032565.

Collins, T.M., O. Fedrigo and G.J. Naylor. 2005. Choosing the best genes for the job: the case for stationary genes in genome-scale phylogenetics. Syst. Biol., 54: 493-500. https://doi.org/10.1080/10635150590947339.

Corrigan, S., D.P. Maisano, C. Eddy, C. Duffy, L. Yang, C. Li, A.L. Bazinet, S. Mona and G.J.P. Naylor. 2017. Historical introgression drives pervasive mitochondrial admixture between two species of pelagic sharks. Mol. Phylogenet. Evol., 110: 122-126. https://doi.org/10.1016/j.ympev.2017.03.011.

Davey, J.W. and M.L. Blaxter. 2010. RADSeq: next-generation population genetics. Brief Funct Genomics, 9: 416-423. https://doi.org/10.1093/bfgp/elq031.

Delsuc, F., H. Brinkmann and H. Philippe. 2005. Phylogenomics and the reconstruction of the tree of life. Nat. Rev. Genet., 6: 361-375. https://doi.org/10.1038/nrg1603.

Emerson, K.J., C.R. Merz, J.M. Catchen, P.A. Hohenlohe, W.A. Cresko, W.E. Bradshaw and C.M. Holzapfel. 2010. Resolving post-glacial phylogeography using high-throughput sequencing. Proc. Natl. Acad. Sci. U.S.A., 107: 16196-16200. https://doi.org/10.1073/pnas.1006538107.

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. Syst. Biol., 61: 717-726. https://doi.org/10.1093/sysbio/sys004.

Funk, D.J. and K.E. Omland. 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annu. Rev. Ecol. Evol. Syst., 34: 397-423. https://doi.org/10.1146/annurev.ecolsys.34.011802.132421.

Hebert, F.O., S. Renaut and L. Bernatchez. 2013. Targeted sequence capture and resequencing implies a predominant role of regulatory regions in the divergence of a sympatric lake whitefish species pair (*Coregonus clupeaformis*). Mol. Ecol., 22: 4896-4914. https://doi.org/10.1111/mec.12447.

Hedtke, S.M., M.J. Morgan, D.C. Cannatella and D.M. Hillis. 2013. Targeted enrichment: maximizing orthologous gene comparisons across deep evolutionary time. PLoS ONE, 8: e67908. https://doi.org/10.1371/journal.pone.0067908.

Huang, J.M., H. Yuan and C.H. Li. 2021. Protocol for cross-species target-gene enrichment. Bio-101: e1010606. https://doi.org/10.21769/BioProtoc.1010606.

Hugall, A.F., T.D. O'Hara, S. Hunjan, R. Nilsen and A. Moussalli. 2016. An exon-capture system for the entire class Ophiuroidea. Mol. Biol. Evol., 33: 281-294. https://doi.org/10.1093/molbev/msv216.

Hughes, L.C., G. Ortí, Y. Huang, Y. Sun, C.C. Baldwin, A.W. Thompson, D. Arcila, R.R. Betancur, C. Li, L. Becker, N. Bellora, X. Zhao, X. Li, M. Wang, C. Fang, B. Xie, Z. Zhou, H. Huang, S. Chen, B. Venkatesh and Q. Shi. 2018. Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. Proc. Natl. Acad. Sci. U.S.A., 115: 6249-6254. https://doi.org/10.1073/pnas.1719358115.

Hughes, L.C., G. Ortí, H. Saad, C. Li, W.T. White, C.C. Baldwin, K.A. Crandall, D. Arcila and R.R. Betancur. 2021. Exon probe sets and bioinformatics pipelines for all levels of fish phylogenomics. Mol. Ecol. Resour., 21: 816-833. https://doi.org/10.1111/1755-0998.13287.

Ilves, K.L. and H. López-Fernández. 2014. A targeted next-generation sequencing toolkit for exon-based cichlid phylogenomics. Mol. Ecol. Resour., 14: 802-811. https://doi.org/10.1111/1755-0998.12222.

Ilves, K.L., D. Torti and H. López-Fernández. 2018. Exon-based phylogenomics strengthens the phylogeny of Neotropical cichlids and identifies remaining conflicting clades (Cichliformes: Cichlidae: Cichlinae). Mol. Phylogenet. Evol., 118: 232-243. https://doi.org/10.1016/j.ympev.2017.10.008.

Jarvis, E.D., S. Mirarab, A.J. Aberer, B. Li, P. Houde, C. Li, S.Y. Ho, B.C. Faircloth, B. Nabholz, J.T. Howard, A. Suh, C.C. Weber, R.R. da Fonseca, J. Li, F. Zhang, H. Li, L. Zhou, N. Narula, L. Liu, G. Ganapathy, B. Boussau, M.S. Bayzid, V. Zavidovych, S. Subramanian, T. Gabaldón, S. Capella-Gutiérrez, J. Huerta-Cepas, B. Rekepalli, K. Munch, M. Schierup, B. Lindow, W.C. Warren, D. Ray, R.E. Green, M.W. Bruford, X. Zhan, A. Dixon, S. Li, N. Li, Y. Huang, E.P. Derryberry, M.F. Bertelsen, F.H. Sheldon, R.T. Brumfield, C.V. Mello, P.V. Lovell, M. Wirthlin, M.P. Schneider, F. Prosdocimi, J.A. Samaniego, A.M. Vargas Velazquez, A. Alfaro-Núñez, P.F. Campos, B. Petersen, T. Sicheritz-Ponten, A. Pas, T. Bailey, P. Scofield, M. Bunce, D.M. Lambert, Q. Zhou, P. Perelman, A.C. Driskell, B. Shapiro, Z. Xiong, Y. Zeng, S. Liu, Z. Li, B. Liu, K. Wu, J. Xiao, X. Yinqi, Q. Zheng, Y. Zhang, H. Yang, J. Wang, L. Smeds, F.E. Rheindt, M. Braun, J. Fjeldsa, L. Orlando, F.K. Barker, K.A. Jønsson, W. Johnson, K.P. Koepfli, S. O'Brien, D. Haussler, O.A. Ryder, C. Rahbek, E. Willerslev, G.R. Graves, T.C. Glenn, J. McCormack, D. Burt, H. Ellegren, P. Alström, S.V. Edwards, A. Stamatakis, D.P. Mindell, J. Cracraft, E.L. Braun, T. Warnow, W. Jun, M.T. Gilbert and G. Zhang. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science, 346: 1320-1331. https://doi.org/10.1126/science.1253451.

Jensen, M.R., E.E. Sigsgaard, S. Liu, A. Manica, S.S. Bach, M.M. Hansen, P.R. Møller and P.F. Thomsen. 2021. Genome-scale target capture of mitochondrial and nuclear environmental DNA from water samples. Mol. Ecol. Resour., 21: 690-702. https://doi.org/10.1111/1755-0998.13293.

Jiang, J., H. Yuan, X. Zheng, Q. Wang, T. Kuang, J. Li, J. Liu, S. Song, W. Wang, F. Cheng, H. Li, J. Huang and C. Li. 2019. Gene markers for exon capture and phylogenomics in ray-finned fishes. Ecol. Evol., 9: 3973-3983. https://doi.org/10.1002/ece3.5026.

Kolmann, M.A., L.C. Hughes, L.P. Hernandez, D. Arcila, R.R. Betancur, M.H. Sabaj, H. López-Fernández and G. Ortí. 2021. Phylogenomics of piranhas and pacus (Serrasalmidae) uncovers how dietary convergence and parallelism obfuscate traditional morphological taxonomy. Syst. Biol., 70: 576-592. https://doi.org/10.1093/sysbio/syaa104.

Kuang, T., L. Tornabene, J. Li, J. Jiang, P. Chakrabarty, J.S. Sparks, G.J.P. Naylor and C. Li. 2018. Phylogenomic analysis on the exceptionally diverse fish clade Gobioidei (Actinopterygii: Gobiiformes) and data-filtering based on molecular clocklikeness. Mol. Phylogenet. Evol., 128: 192-202. https://doi.org/10.1016/j.ympev.2018.07.018.

Kurland, C.G., B. Canback and O.G. Berg. 2003. Horizontal gene transfer: a critical view. Proc. Natl. Acad. Sci. U.S.A., 100: 9658-9662. https://doi.org/10.1073/pnas.1632870100.

Lappin, F.M., R.L. Shaw and D.J. Macqueen. 2016. Targeted sequencing for high-resolution evolutionary analyses following genome duplication in salmonid fish: Proof of concept for key components of the insulin-like growth factor axis. Mar. Genomics, 30: 15-26. https://doi.org/10.1016/j.margen.2016.003.

Lemmon, A.R., S.A. Emme and E.M. Lemmon. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst. Biol., 61: 727-744. https://doi.org/10.1093/sysbio/sys049.

Li, C. 2007. A genome-scale approach to phylogeny of ray-finned fish (Actinopterygii) and molecular systematics of Clupeiformes. School of Biological Sciences. Doctoral Thesis, University of Nebraska, Lincoln, Nebraska, U.S.A., 118pp.

Li, C., S. Corrigan, L. Yang, N. Straube, M. Harris, M. Hofreiter, W.T. White and G.J. Naylor. 2015. DNA capture reveals transoceanic gene flow in endangered river sharks. Proc. Natl. Acad. Sci. U.S.A., 112: 13302-13307. https://doi.org/10.1073/pnas.1508735112.

Li, C., M. Hofreiter, N. Straube, S. Corrigan and G. Naylor. 2013. Capturing protein-coding genes across highly divergent species. BioTechniques, 54: 321-326. https://doi.org/10.2144/000114039.

Li, C., K.A. Matthes-Rosana, M. Garcia and G.J. Naylor. 2012. Phylogenetics of Chondrichthyes and the problem of rooting phylogenies with distant outgroups. Mol. Phylogenet. Evol., 63: 365-373. https://doi.org/10.1016/j.ympev.2012.01.013.

Li, H., Y. He, J. Jiang, Z. Liu and C. Li. 2018. Molecular systematics and phylogenetic analysis of the Asian endemic freshwater sleepers (Gobiiformes: Odontobutidae). Mol. Phylogenet. Evol., 121: 1-11. https://doi.org/10.1016/j.ympev.2017.12.026.

Liu, J., J. Jiang, S. Song, L. Tornabene, R. Chabarria, G.J. Naylor and C. Li. 2017. Multilocus DNA barcoding-species identification with multilocus data. Sci. Rep., 7: 1-12. https://doi.org/10.1038/s41598-017-16920-2.

Luikart, G., P.R. England, D. Tallmon, S. Jordan and P. Taberlet. 2003. The power and promise of population genomics: from genotyping to genome typing. Nat. Rev. Genet., 4: 981-994. https://doi.org/10.1038/nrg1226.

Maddison, W.P. 1997. Gene trees in species trees. Syst. Biol., 46: 523-536. https://doi.org/10.1093/sysbio/46.3.523.

Maddison, W.P. and L.L. Knowles. 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol., 55: 21-30. https://doi.org/10.1080/10635150500354928.

Maisano Delser, P., S. Corrigan, M. Hale, C. Li, M. Veuille, S. Planes, G. Naylor and S. Mona. 2016. Population genomics of *C. melanopterus* using target gene capture data: demographic inferences and conservation perspectives. Sci. Rep., 6: 33753. https://doi.org/10.1038/srep33753.

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

Nielsen, E.E., J.A.T. Morgan, S.L. Maher, J. Edson, M. Gauthier, J. Pepperell, B.J. Holmes, M.B. Bennett and J.R. Ovenden. 2017. Extracting DNA from 'jaws': high yield and quality from archived tiger shark (*Galeocerdo cuvier*) skeletal material.

Mol. Ecol. Resour., 17: 431-442. https://doi.org/10.1111/1755-0998.12580.

Peterson, B.K., J.N. Weber, E.H. Kay, H.S. Fisher and H.E. Hoekstra. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE, 7: e37135. https://doi.org/10.1371/journal.pone.0037135.

Rincon-Sandoval, M., R.R. Betancur and J.A. Maldonado-Ocampo. 2019. Comparative phylogeography of trans-Andean freshwater fishes based on genome-wide nuclear and mitochondrial markers. Mol. Ecol., 28: 1096-1115. https://doi.org/10.1111/mec.15036.

Roa-Varón, A., R.B. Dikow, G. Carnevale, L. Tornabene, C.C. Baldwin, C. Li and E.J. Hilton. 2021. Confronting sources of systematic error to resolve historically contentious relationships: a case study using Gadiform fishes (Teleostei, Paracanthopterygii, Gadiformes). Syst. Biol., 70: 739-755. https://doi.org/10.1093/sysbio/syaa095.

Sarker, A., J. Jiang, H. Naher, J. Huang, K.K. Sarker, G. Yin, M.A. Baki and C. Li. 2021. Cross-species gene enrichment revealed a single population of Hilsa shad (*Tenualosa ilisha*) with low genetic variation in Bangladesh waters. Sci. Rep., 11: 11560. https://doi.org/10.1038/s41598-021-90864-6.

Sarker, A., H. Naher, J. Huang, K.K. Sarker, M.A. Baki and C. Li. 2020. Genetic diversity of *Hilsa kelee* collected from the Bay of Bengal and the Arabian Sea. Mar. Biodivers., 50: 94. https://doi.org/10.1007/s12526-020-01114-3.

Song, S., J. Zhao and C. Li. 2017. Species delimitation and phylogenetic reconstruction of the sinipercids (Perciformes: Sinipercidae) based on target enrichment of thousands of nuclear coding sequences. Mol. Phylogenet. Evol., 111: 44-55. https://doi.org/10.1016/j.ympev.2017.03.014.

Straube, N., C. Li, J.M. Claes, S. Corrigan and G.J. Naylor. 2015. Molecular phylogeny of Squaliformes and first occurrence of bioluminescence in sharks. BMC Evol. Biol., 15: 162. https://doi.org/10.1186/s12862-015-0446-6.

Straube, N., C. Li, M. Mertzen, H. Yuan and T. Moritz. 2018. A phylogenomic approach to reconstruct interrelationships of main clupeocephalan lineages with a critical discussion of morphological apomorphies. BMC Evol. Biol., 18: 158. https://doi.org/10.1186/s12862-018-1267-1.

Therkildsen, N.O. and S.R. Palumbi. 2017. Practical low-coverage genomewide sequencing of hundreds of individually barcoded samples for population and evolutionary genomics in non-model species. Mol. Ecol. Resour., 17: 194-208. https://doi.org/10.1111/1755-0998.12593.

White, W.T., S. Corrigan, L. Yang, A.C. Henderson, A.L. Bazinet, D.L. Swofford and G.J.P. Naylor. 2018. Phylogeny of the manta and devilrays (Chondrichthyes: Mobulidae), with an updated taxonomic arrangement for the family. Zool. J. Linn. Soc., 182: 50-75. https://doi.org/10.1093/zoolinnean/zlx018.

Yin, G., Y. Pan, A. Sarker, M.A. Baki, J.K. Kim, H. Wu and C. Li. 2019. Molecular systematics of *Pampus* (Perciformes: Stromateidae) based on thousands of nuclear loci using target-gene enrichment. Mol. Phylogenet. Evol., 140: 106595. https://doi.org/10.1016/j.ympev.2019.106595.

Zong, C., S. Lu, A.R. Chapman and X.S. Xie. 2012. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science, 338: 1622-1626. https://doi.org/10.1126/science.1229164.

엑손 포획 - 원리와 어류의 계통유전체학 및 집단유전체학으로의 응용

Chenhong Li[1,2]

[1]상하이해양대학교 해양동물 분류 및 진화 핵심연구소, [2]상하이해양대학교 수생동물 유전 및 육종 협력 혁신센터

요 약 : 한 유전좌위 또는 소수의 유전좌위에 기반한 계통발생학적 재구성은 분자 계통수/종 계통수의 불일치로 인해 오해를 불러일으킬 수 있다. 종의 구분과 종내 연구에서도 적은 유전좌위를 사용할 때 통계력 부족으로 해상도가 낮은 경우가 많이 발생한다. 엑손 포획법은 게놈 규모의 데이터를 수집하는 가장 효율적인 방법 중 하나로, 종내 및 상위 수준에서 생물의 패턴과 역사를 구명하는 연구에 크게 이바지할 수 있다. 이 논문에서는 단일 유전자 방법에서 게놈 접근으로의 전환의 진보와 게놈 기술에 비해 엑손 포획법의 적용 이점을 설명하였다. 또한 엑손 포획법의 원리를 설명하고 이 방법의 적용을 위한 상세한 제언을 기술하였다. 최종적으로, 두 가지 적용을 활용한 엑손 포획법을 설명하고 이 기술에 대한 미래 전망을 논의하였다.

찾아보기 낱말 : 표적농축, 엑손 포획, 종 경계, 집단 구조, 개체군 동태, 적응, 분자계통분류학, 계통발생학