

Applications of Next-Generation Sequencing Technologies and Computational Tools in Molecular Evolution and Aquatic Animals Conservation Studies: A Short Review

Evolutionary Bioinformatics
Volume 15: 1–5
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1176934319892284



Min Pau Tan^{1,2} , Li Lian Wong^{1,3}, Siti Aisyah Razali¹,
Nor Afiqah-Aleng¹, Siti Azizah Mohd Nor¹, Yeong Yik Sung¹,
Yves Van de Peer^{1,4,5}, Patrick Sorgeloos^{1,6}
and Muhd Danish-Daniel^{1,2} 

¹Institute of Marine Biotechnology, Universiti Malaysia Terengganu, Kuala Nerus, Malaysia.

²Faculty of Fisheries and Food Science, Universiti Malaysia Terengganu, Kuala Nerus, Malaysia.

³Institute of Tropical Aquaculture, Universiti Malaysia Terengganu, Kuala Nerus, Malaysia. ⁴Center for Plant Systems Biology, VIB, Ghent, Belgium. ⁵Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium. ⁶Laboratory of Aquaculture & Artemia Reference Center, Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium.

ABSTRACT: Aquatic ecosystems that form major biodiversity hotspots are critically threatened due to environmental and anthropogenic stressors. We believe that, in this genomic era, computational methods can be applied to promote aquatic biodiversity conservation by addressing questions related to the evolutionary history of aquatic organisms at the molecular level. However, huge amounts of genomics data generated can only be discerned through the use of bioinformatics. Here, we examine the applications of next-generation sequencing technologies and bioinformatics tools to study the molecular evolution of aquatic animals and discuss the current challenges and future perspectives of using bioinformatics toward aquatic animal conservation efforts.

KEYWORDS: Bioinformatics, molecular evolution, aquatic animals, conservation, genomics

RECEIVED: November 7, 2019. **ACCEPTED:** November 12, 2019.

TYPE: Molecular Evolution - Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Strategic Research Grant vot 55195 and Universiti Malaysia Terengganu internal funds to the Bioinformatics Program.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Muhd Danish-Daniel, Institute of Marine Biotechnology, Universiti Malaysia Terengganu, Kuala Nerus 21030, Malaysia.
Email: mdda@umt.edu.my

Introduction

Approximately 70% of the Earth's surface is covered with water. Within this aquatic environment lives a whole myriad of aquatic organisms. These life forms are vital to the global ecosystem sustainability. From an evolutionary point of view, many terrestrial organisms are said to be of aquatic origin and have evolved through millions of years.¹ In other words, it is crucial to study the evolutionary history of aquatic organisms, many of which remain taxonomically challenging to correctly identify, to better understand the origin of all life forms on Earth.

The rapid growth of human populations and development—in particular, land reclamation of coastal regions, harvesting of fishery resources, and aquaculture practices—has caused much public concern about their negative impacts on the environment.² Master et al³ reported that the aquatic fauna of the United States is at high risk of extinction, where up to 70% of all freshwater mussels, 49% of freshwater fishes, 30% of plants, and 20% of mammals and birds are endangered. Global rates are similar for those groups.⁴ Many studies have demonstrated that exotic species, habitat loss, pollution, and unsustainable exploitation accounted for most of the extinctions of marine's wildlife.⁵

Putting aside climate change, this ongoing “biodiversity decline” is truly a catastrophe to all species that fully relies on

this biosphere to live in. Scientists all around the world have been working on trying to conserve aquatic biodiversity, and one of the tools they have at hand to help in this endeavor is bioinformatics. Indeed, bioinformatics tools can be very useful in studies where the data provide better understanding on the mechanistic underlying the evolution of organisms at the molecular level and can be very helpful for designing holistic conservation and management strategies. For instance, the evolutionary processes that are at play when a species colonizes a new environment provide an opportunity to explore the mechanisms underlying genetic adaptation, which is an essential knowledge for understanding evolution and the maintenance of biodiversity.⁶

One of the most exciting technological advancements during the past decades was the development of powerful and high-throughput nucleic acid sequencing techniques to solve questions in phylogenetics and molecular evolution of taxa or their complexes.⁷ With the advent of these sophisticated molecular techniques, huge amounts of data have become available. Comprehending these massive amounts of data requires advanced bioinformatics skills and databases to collate. Furthermore, as these data are stored at numerous databases, both public and private institution-based, there is a need to link the different databases to conduct an exhaustive analysis of the data.



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<http://www.creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Knowledge about the evolutionary relationships among species has been used in many important biotechnological applications. For example, the understanding of viral quasi-species variation allows us to trace routes of infectious disease transmission.⁸ The analysis of the host-pathogen relationships in terms of mutual genetic variation can lead to deeper insights into drug design for medical and agricultural purposes. Structural biologists are now focusing on the phylogeny of related organisms to study sets of homologous proteins because these reflect different variants stored in nature and which can reveal structural and functional constraints.⁹ While huge amounts of data can now relatively easily and cheaply be produced through, for instance, next-generation sequencing (NGS) techniques, understanding the underlying principles for its applications could only be disentangled through the application of bioinformatics tools.

In this article, we aim to provide insights into the power and usefulness of bioinformatics to better understand molecular evolutionary processes in aquatic animals and how this information can be used as an important input for their conservation. First, we will briefly discuss how the contribution of knowledge of molecular evolution can be important toward aquatic animal conservation followed by some insights on the bioinformatics tools used for this purpose. At the end, we also share some of our thoughts on the challenges and future perspectives of bioinformatics in conservation studies.

Importance of Species Identification Techniques as the First Step Toward Aquatic Animals Conservation

Each aquatic organism has its own unique heredity which makes it special. Each successful species has had the ability to survive in a changing and challenging environment, through adapting to changes, developing immunity to disease, and through selective fitness over generations.¹⁰ Sustaining biodiversity is critical in maintaining the health of our environment and improving the quality of human life. Conserving aquatic animals, plants, and algae will provide food for the growing human population, increases oxygen, and reduces carbon dioxide in the atmosphere; facilitate drug discovery; and has numerous other downstream applications.¹¹ Every living organism has an important role to play in the ecosystem, either independently or in close interaction with the environment, and thus each has its own value. Without a rich biodiversity, we would have lower food security, limited supply of pharmaceutical drugs, a less healthy environment, and poorer economic status.

Accurate species identification is the basis for addressing many molecular ecological questions and is fundamental to management and conservation. While morphology is often the most economical approach to species identification, there are many circumstances that molecular-based techniques may be particularly useful especially when dealing with cryptic species, juveniles, incomplete specimens, hybrids, and new, unknown species. For species identification, DNA barcoding was and is a taxonomic tool that uses a short, standardized region of the

mitochondrial DNA to identify organisms to the species level. Nucleotide sequences for the selected marker of an unknown specimen are then aligned and edited, for example, with a tool such as ClustalW implemented in the Molecular Evolutionary Genetics Analysis (MEGA) software package.^{12,13}

The sequences then are compared with a public database, developed from input of researchers worldwide. The accuracy of species identification relies wholly on the data deposited and the discriminatory power of the molecular markers used. For instance, the DNA marker commonly used for species identification in barcoding fishes is a 648-bp (base pair) region of the mitochondrial DNA called cytochrome c oxidase subunit I (*col*). The barcoding reference data are stored and managed in online global databases (eg, <http://www.boldsystems.org>).

Recent technologies have provided more advanced and sensitive tools for species detection using the genetic material that is present in the environment, thus so-called environmental DNA (eDNA).¹⁴ Environmental DNA was initially applied to bacterial community composition and functional diversity studies, but recently scientists have started to use eDNA on macrofaunal studies to monitor the presence/absence of rare, endangered, indicator, and invasive species through environmental samples such as water and soil samples, without the need for direct sampling of the target species.¹⁵⁻¹⁸ There are shotgun metagenomics and metabarcoding methods that can be used to study the eDNA. While both approaches involve NGS of DNA, the methods serve different purposes and the selection of method relies on the research question being addressed. In shotgun metagenomics, one sequences the total eDNA present in the sample to understand the community composition and functional diversity, whereas metabarcoding uses one or more barcoding genes to detect the presence of a targeted taxonomic group from the soil, water, or air samples, to understand the biodiversity and its abundance. Metabarcoding relies on polymerase chain reaction amplification of gene fragments using a given primer set or sets. Metabarcoding of eDNA has proven to be reliable and cost-effective for monitoring of fish,¹⁹ fish and amphibian,²⁰ and pathogens in aquaculture.^{21,22}

Some Selected Molecular, Computational, and Bioinformatics Tools

Conservation genetics is an applied science employing molecular tools to study the genetic structure, evolutionary patterns, and interaction process within the context of biodiversity conservation.²³ Neutral molecular markers such as random fragment length polymorphism, amplified fragment length polymorphism, random amplification of polymorphic DNA, single-strand conformation polymorphism, minisatellites, microsatellites, single-nucleotide polymorphisms (SNPs) are used in many conservation genetic studies to unravel the importance of genetic data for taxonomic distinction and management of conservation units. The usage of a small number of neutral markers has raised debatable reliability of these markers in representing the population and species variations at the

level of an entire genome, given that these neutral markers are not subjected to selection and local adaptation.

These limitations and the considerable reduction in the cost of NGS have further pushed forward the needs for the transition from conservation genetics to conservation genomics. The application of NGS increases the estimation accuracy for genetic variation at finer population scales, when genome-wide screening of thousands of markers is conducted. These studies are especially useful for marine aquatic species with high dispersion rates allowing gene flow regardless of geographical distance.^{24,25} Next-generation sequencing facilitates the study of gene interaction with environmental changes as it could determine variations in both neutral and nonneutral markers. Single-nucleotide polymorphism arrays generated by NGS studies have also revealed that putatively adaptive markers provide stronger differentiation signals compared with neutral markers, given that the former strictly relies on selective forces for estimating population divergence.^{26,27} Next-generation sequencing also allows for the study of gene expression. Expression patterns of various genes in host aquatic species corresponding to multiple environmental stressors or inducers have been well documented.²⁸⁻³²

Currently used NGS platforms include Illumina (formerly Solexa) sequencing, SOLiD system of ABI, the Polonator G.007, Helicos Heliscope, PacBio SMRT sequencing, and Oxford Nanopore. The choice of technologies depends on the throughput capacity, running time, coverage depth, simultaneous multiplexing, cost, and error rates. Experimental designs and sequencing strategies to be applied should fully depend on 3 main categories of research questions,³³ namely, (1) genome-wide screening of genetic variation, (2) identifying nonneutral variation, and (3) integrating environmental and genetic parameters with gene expression analysis. One of the important keys to all conservation genomics studies is the detection and screening of SNPs within the genome, and there are various approaches available with their specific final goals and available resources. Whole genome sequencing and transcriptome sequencing would be great choices for development of SNPs for many follow-up experiments.³³ However, if only a single experiment is performed for a population screening of a nonmodel species, SNPs could be identified using RAD-Tag sequencing.³³ The SNPs can then be screened for either with SNP-chip or with a RAD-tag sequencing procedure. To address the second research question, which is to identify markers involved in adaptation and screening of population for variation in these markers, methods based on NGS such as genome-wide selection scans, genome-wide association studies, and gene-environment association studies can be applied. These methods allow the revelation of genes associated with selection and adaptation of the physiological mechanisms to the changing aquatic environment.³⁴⁻³⁶ Finally, to identify genes that are associated with populations with different genetic heritability or environmental quality, transcriptome analysis using RNA-seq procedure can be performed.^{37,38}

Applications of NGS involve the management of massive data sets requiring a huge data storage facility and bioinformatics pipelines to effectively compile, process, and analyze the sequence data. An extensive list of bioinformatics tools with respective functions and usages for downstream population genomic analyses is available. Identification of genotyping errors and data filtering programs is crucial in improving data quality. Erroneous data in SNP data sets can be assessed by performing simple estimation of departure from Hardy-Weinberg equilibrium (HWE), and probabilistic genotype calling programs such as ANGSD,³⁹ and ngsCovar.⁴⁰ Data set filtering could highly affect downstream summary statistics. Filtering using minor allele frequencies thresholds is one of the most common methods in filtering RAD-seq data sets, which allows for the removal of sequencing errors and rare alleles. However, to produce robust genetic and demographic inferences, running a trial on filtering parameters is crucial.⁴¹ Downstream computational analyses include the estimation and measurement of various parameters depending on the research question asked. In the context of genomic analysis, most commonly analyzed parameters for aquatic species inbreeding events are multiple-locus heterozygosity and genomic relatedness matrices using Arlequin ver. 3.5.1.2.⁴² Using the “adegenet” package,⁴³ Mantel test estimates recent demographic history and the correlation between population divergences with geographical distance in populations of aquatic species. Other bioinformatics programs to infer complex demographic models and population clustering patterns include multivariate analyses such as principal component analysis and discriminate analysis of principal components using adegenet R and non-model-based methods such as STRUCTURE and ADMIXTURE.^{44,45} These clustering analysis programs assign individual aquatic species to their respective populations of origin. Comprehensive analysis and interpretation of data based on reliable computational methods are important in producing robust and reliable population genomics data that can then be applied to evolutionary biology and biodiversity conservation.

Current Challenges and Future Perspectives

To date, computational methods and NGS have been employed by conservation and evolutionary geneticists to improve conservation management of marine populations, especially for endangered species.⁴⁶⁻⁴⁸ Despite the technological advancements, analyzing genome-wide molecular data poses a major challenge. Handling large-scale and complex data requires high competency in bioinformatics and the ability to analyze and interpret the vast amount of data and translate them into biological applications.⁴⁹ This is further exacerbated by the rapid development of new bioinformatics tools.⁵⁰ Nevertheless, there are many workshops, seminars, and online courses and materials that are available to assist researchers in this field.^{49,50}

Next-generation sequencing technologies facilitate the sequencing process with advanced sequencing length and accuracy but genome assembly remains a significant challenge.

Current challenges in building a genome assembler include considering sequencing errors, the high-throughput nature of sequencers, short-read lengths, and genomic repeats.^{51–53} However, these problems can, at least partly, be resolved by using the newest third-generation sequencing platforms, improved data analysis, and advanced mapping technologies and sequence assembly algorithms.

Other challenges in genomic analysis of aquatic species is the more limited number of available reference genomes. Compared with terrestrial species, genomic studies on aquatic species are still lagging behind.⁵⁰ For instance, the variable pattern of the genomic architecture of aquatic animals is difficult to observe due to limited information about the genomic location of assayed markers. However, knowledge about the genomic location of assayed markers could be achieved not only through access to full-genome reference sequences of the target species but also through those of closely related species.⁵⁴ Thus, although a limited number of high-quality genomes for new species is a challenge to the researchers, it is still possible to extract useful information from other available genomics resources.^{55–57} The rapid emergence of aquatic resources will facilitate future genomics studies in a broader set of related species, which will provide a better understanding in basic evolutionary and conservation processes in aquatic species.

Upstream process of pipeline is also one of the major challenges in aquatic genomic analysis, ie, the conditions and criteria used for sample preparation. This step is crucial in ensuring high-quality downstream bioinformatics analysis.⁵⁸ Aquatic animals have a broad range of qualitative class and criteria, which have to be taken into consideration prior to proceeding with the sequencing process. Biological diversity, nonindigenous species, exploited aquatic animals, and contaminants are the examples of qualitative descriptors and conditions that might affect the data generation.⁵⁹ Major efforts need to be performed in the future to build biodiversity monitoring and research infrastructures (data generation, data storage and curation, and data analysis) in addressing this issue.

Conclusions

In conclusion, conservation of aquatic species is vital in ensuring the sustainability of biodiversity. In this era of genomics, bioinformatics and the application of computational and statistical tools play a major role in elucidating the evolutionary processes of aquatic organisms at the molecular level. Ultimately, these tools can provide important indicators for implementing conservation strategies. Despite the still many technological challenges in applying bioinformatics as a conservation approach, all stakeholders must actively involve in making this effort a reality.

Acknowledgements

This article was prepared as part of the efforts to develop a bioinformatics research program at the Institute of Marine


Biotechnology, UMT. We thank all the people involved directly or indirectly in the preparation of this manuscript and special thanks to UMT for allocating internal funding to the Bioinformatics Program and UMT Center for Research and Innovation Management (CRIM) for providing the Strategic Research Grant under vot no. 55195.

Author Contributions

MPT and MDD designed the review outline; MPT, MDD, LLW, SAR, and NAA wrote the manuscript; SAMN, YYS, YVDP, and PS revised the manuscript. All authors read and approved the final manuscript.

ORCID iDs

Min Pau Tan  <https://orcid.org/0000-0003-0149-5247>

Muhd Danish-Daniel  <https://orcid.org/0000-0001-9438-0128>

REFERENCES

1. Delwiche CF, Cooper ED. The evolutionary origin of a terrestrial flora. *Curr Biol*. 2015;25:R899–R910.
2. Diana JS. Aquaculture production and biodiversity conservation. *Bioscience*. 2009;59:27–38.
3. Master LL, Flack SR, Stein BA (eds). *Rivers of Life: Critical Watersheds for Protecting Freshwater Biodiversity*. Arlington, VA: Nature Conservancy; 1998.
4. Millennium Ecosystem Assessment (MEA). *Ecosystems and Human Well-Being: Biodiversity Synthesis*. Washington, DC: World Resources Institute; 2005.
5. McClenachan L, Cooper AB, Carpenter KE, Dulvy NK. Extinction risk and bottlenecks in the conservation of charismatic marine species. *Conservation Letters*. 2012;5:73–80.
6. Hill J, Enbody ED, Pettersson ME, et al. Recurrent convergent evolution at amino acid residue 261 in fish rhodopsin. *Proc Natl Acad Sci U S A*. 2019;116:18473–18478.
7. Patwardhan A, Ray S, Roy A. Molecular markers in phylogenetic studies—a review. *J Phylogenetics Evolut Biol*. 2014;2:131.
8. Skums P, Zelikovsky A, Singh R, et al. QUENTIN: reconstruction of disease transmissions from viral quasispecies genomic data. *Bioinformatics*. 2017;34:163–170.
9. Lio P, Goldman N. Models of molecular evolution and phylogeny. *Genome Res*. 1998;8:1233–1244.
10. Hamilton PB, Rolshausen G, Uren Webster TM, Tyler CR. Adaptive capabilities and fitness consequences associated with pollution exposure in fish. *Philos Trans R Soc Lond B Biol Sci*. 2017;372:20160042.
11. Alho CJ. The importance of biodiversity to human health: an ecological perspective. *Estud Av*. 2012;26:151–166.
12. Kumar S, Tamura K, Nei M. *MEGA: Molecular Evolutionary Genetics Analysis. Version 1.0*. University Park, PA: The Pennsylvania State University; 1993.
13. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol Biol Evol*. 2018;35:1547–1549.
14. Taberlet P, Bonin A, Zinger L, Coissac E. *Environmental DNA: For Biodiversity Research and Monitoring*. Oxford, UK: Oxford University Press; 2018.
15. Ficetola GF, Miaud C, Pompanon F, Taberlet P. Species detection using environmental DNA from water samples. *Biol Lett*. 2008;4:423–425.
16. Rees H, Maddison B, Middleditch D, Patmore J, Gough K. The detection of aquatic animal species using environmental DNA—a review of eDNA as a survey tool in ecology. *J Appl Ecol*. 2014;51:1450–1459.
17. Clusa L, Miralles Basanta A, Escot C, Garcia-Vazquez E. eDNA for detection of five highly invasive molluscs. A case study in urban rivers from the Iberian Peninsula. *PLoS ONE*. 2017;12:e0188126.
18. Klymus KE, Marshall NT, Stepien CA. Environmental DNA (eDNA) metabarcoding assays to detect invasive invertebrate species in the Great Lakes. *PLoS ONE*. 2017;12:e0177643.
19. Pont D, Rocle M, Valentini A, Civade R, Jean P, Maire A. Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers despite its downstream transportation. *Scientific Reports*. 2018;8:10361.
20. Valentini A, Taberlet P, Miaud C, et al. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Mol Ecol*. 2016;25:929–942.

21. Gomes GB, Hutson KS, Domingos JA, et al. Use of environmental DNA (eDNA) and water quality data to predict protozoan parasites outbreaks in fish farms. *Aquaculture*. 2017;479:467-473.
22. Peters L, Spatharis S, Dario MA, et al. Environmental DNA: a new low-cost monitoring tool for pathogens in salmonid aquaculture. *Front Microbiol*. 2018;9:3009.
23. Avise JC. Perspective: conservation genetics enters the genomics era. *Conserv Genet*. 2010;11:665-669.
24. Hollenbeck CM, Portnoy DS, Gold JR. Evolution of population structure in an estuarine-dependent marine fish. *Ecol Evol*. 2019;9:3141-3152.
25. Hemmer-Hansen J, Hüsey K, Baktoft H, et al. Genetic analyses reveal complex dynamics within a marine fish management area. *Evol Appl*. 2019;12:830-844.
26. Salas EM, Bernardi G, Berumen ML, Gaither MR, Rocha LA. RADseq analyses reveal concordant Indian Ocean biogeographic and phylogeographic boundaries in the reef fish *Dascyllus trimaculatus*. *R Soc Open Sci*. 2019;6:172413.
27. Carreras C, Ordóñez V, Zane L, et al. Population genomics of an endemic Mediterranean fish: differentiation by fine scale dispersal and adaptation. *Sci Rep*. 2017;7:43417.
28. Lacaze É, Gendron AD, Miller JL, et al. Cumulative effects of municipal effluent and parasite infection in yellow perch: a field study using high-throughput RNA-sequencing. *Sci Total Environ*. 2019;665:797-809.
29. Shi KP, Dong SL, Zhou YG, Li Y, Gao QF, Sun DJ. RNA-seq reveals temporal differences in the transcriptome response to acute heat stress in the Atlantic salmon (*Salmo salar*). *Comp Biochem Physiol Part D Genomics Proteomics*. 2019;30:169-178.
30. Bian C, Huang Y, Li J, et al. Divergence, evolution and adaptation in ray-finned fish genomes. *Sci China Life Sci*. 2019;62:1003-1018.
31. Jones J, Wellband K, Zielinski B, Heath DD. Transcriptional basis of copper-induced olfactory impairment in the sea lamprey, a primitive invasive fish. *G3 (Bethesda)*. 2019;9:933-941.
32. Hart JC, Ellis NA, Eisen MB, Miller CT. Convergent evolution of gene expression in two high-toothed stickleback populations. *PLoS Genet*. 2018;14:e1007443.
33. Angeloni F, Wagemaker N, Vergeer P, Ouborg J. Genomic toolboxes for conservation biologists. *Evol Appl*. 2012;5:130-143.
34. Brennan RS, Healy TM, Bryant HJ, Van La M, Schulte PM, Whitehead A. Genome-wide selection scans integrated with association mapping reveal mechanisms of physiological adaptation across a salinity gradient in killifish. *BioRxiv*. 2018;254854.
35. Lima TG, Burton RS, Willett CS. Genomic scans reveal multiple mito-nuclear incompatibilities in population crosses of the copepod *Tigriopus californicus*. *Evolution*. 2019;73:609-620.
36. Haenel Q, Roesti M, Moser D, MacColl AD, Berner D. Predictable genome-wide sorting of standing genetic variation during parallel adaptation to basic versus acidic environments in stickleback fish. *Evol Lett*. 2019;3:28-42.
37. Marguerat S, Bähler J. RNA-seq: from technology to biology. *Cell Mol Life Sci*. 2010;67:569-579.
38. Montenegro J, Mochida K, Matsui K, et al. Convergent evolution of body color between sympatric freshwater fishes via different visual sensory evolution. *Ecol Evol*. 2019;9:6389-6398.
39. Korneliusen TS, Albrechtsen A, Nielsen R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics*. 2014;15:356.
40. Fumagalli M, Vieira FG, Linderot T, Nielsen R. ngsTools: methods for population genetics analyses from next-generation sequencing data. *Bioinformatics*. 2014;30:1486-1487.
41. Paris JR, Stevens JR, Catchen JM. Lost in parameter space: a road map for stacks. *Methods Ecol Evol*. 2017;8:1360-1373.
42. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. 2010;10:564-567.
43. Jombart T. ADEGENET: a R package for the multivariate analysis of genetic markers. *Bioinformatics*. 2008;24:1403-1405.
44. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19:1655-1664.
45. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155:945-959.
46. Jorge PH, Mastrochirico-Filho VA, Hata ME, et al. Genetic characterization of the fish *Piaractus brachypomus* by microsatellites derived from transcriptome sequencing. *Front Genet*. 2018;9:46.
47. Khan S, Nabi G, Ullah MW, et al. Overview on the role of advance genomics in conservation biology of endangered species. *Int J Genomics*. 2016;2016:3460416.
48. Paixão RV, Ribolli J, Zaniboni-Filho E. Genetic variation of the endangered neotropical catfish *Steindachneridion scriptum* (Siluriformes: Pimelodidae). *Front Genet*. 2018;9:48.
49. Andrews KR, Luikart G. Recent novel approaches for population genomics data analysis. *Mol Ecol*. 2014;23:1661-1667.
50. Kelley JL, Brown AP, Therikildsen NO, Foote AD. The life aquatic: advances in marine vertebrate genomics. *Nat Rev Genet*. 2016;17:523-534.
51. Abdelrahman H, ElHady M, Alcivar-Warren A, et al. Aquaculture genomics, genetics and breeding in the United States: current status, challenges, and priorities for future research. *BMC Genomics*. 2017;18:191.
52. Fuentes-Pardo AP, Ruzzante DE. Whole-genome sequencing approaches for conservation biology: advantages, limitations and practical recommendations. *Mol Ecol*. 2017;26:5369-5406.
53. Sohn J-I, Nam J-W. The present and future of de novo whole-genome assembly. *Brief Bioinform*. 2018;19:23-40.
54. Hemmer-Hansen J, Therikildsen NO, Pujolar JM. Population genomics of marine fishes: next-generation prospects and challenges. *Biol Bull*. 2014;227: 117-132.
55. Galla SJ, Forsdick NJ, Brown L, et al. Reference genomes from distantly related species can be used for discovery of single nucleotide polymorphisms to inform conservation management. *Genes*. 2019;10:E9.
56. Pimentel JSM, Carmo AO, Rosse IC, et al. High-throughput sequencing strategy for microsatellite genotyping using neotropical fish as a model. *Front Genet*. 2018;9:73.
57. Sember A, Bertollo LAC, Ráb P, et al. Sex chromosome evolution and genomic divergence in the fish *Hoplias malabaricus* (Characiformes, Erythrinidae). *Front Genet*. 2018;9:71.
58. Endrullat C, Glökler J, Franke P, Frohme M. Standardization and quality management in next-generation sequencing. *Appl Transl Genom*. 2016;10:2-9.
59. Bourlat SJ, Borja A, Gilbert J, et al. Genomics in marine monitoring: new opportunities for assessing marine health status. *Mar Pollut Bull*. 2013;74:19-31.