



ELSEVIER

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

The first complete mitochondrial genome data of *Hippocampus kuda* originating from Malaysia

Puteri Nur Syahzanani Jahari^{a,*}, Nur Fatihah Abdul Malik^b,
Mohd Shahir Shamsir^c, M. Thomas P. Gilbert^d, Faezah Mohd
Salleh^a

^a Department of Biosciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia

^b Johor Biotechnology & Biodiversity Corporation (J-Biotech), Level 2, Bio-XCell Malaysia, No. 2, Jalan Bioteknologi 1, SiLC Industrial Park, 79200 Iskandar Puteri, Johor, Malaysia

^c Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, Pagoh Higher Education Hub, 84600 Muar, Johor, Malaysia

^d Section for Evolutionary Genomics, The GLOBE Institute, University of Copenhagen, Øster Farimagsgade 5a, 1353, Copenhagen, Denmark

ARTICLE INFO

Article history:

Received 31 March 2020

Revised 8 May 2020

Accepted 12 May 2020

Available online 21 May 2020

Keywords:

Hippocampus kuda

mitogenome

H. kuda clade

ABSTRACT

The spotted seahorse, *Hippocampus kuda* population is exponentially decreasing globally due to habitat loss contributed by massive coastal urbanization as well as its large exploitation for Chinese herbal medicine. Genomic data would be highly useful to improve biomonitoring of seahorse populations in Malaysia via the usage of non-invasive approaches such as water environmental DNA. Here we report the first complete mitogenome of two *H. kuda* individuals originating from Malaysia, generated using BGISEQ-500RS sequencer. The lengths of both mitogenomes are 16,529bp, consisting of 13 protein-coding genes, 22 transfer RNA genes, two ribosomal RNA genes, and a control region. The overall base composition was 32.46% for A, 29.40% for T, 14.73% for G and 23.41% for C with AT rich features (61.86%). The gene organization of Malaysian *H. kuda* were similar to that of most teleost species. A phylogenetic analysis of the genome against mtDNA data from other *Hippocampus* species showed

* Corresponding author.

E-mail address: puterijahari22@gmail.com (P.N.S. Jahari).

that Malaysian *H. kuda* samples clustered with *H. capensis*, *H. reidi* and *H. kuda*. Notably however, analysis of the data using BLASTn revealed they had 99.18% similarity to *H. capensis*, and only 97.66% to *H. kuda* and *H. reidi*, which are all part of the unresolved *H. kuda* complex. The mitogenomes are deposited in Genbank under the accession number MT221436 (HK1) and MT221436 (HK2).

© 2020 The Author(s). Published by Elsevier Inc.
This is an open access article under the CC BY license.
(<http://creativecommons.org/licenses/by/4.0/>)

Specifications Table

Subject	Genomics
Specific subject area	Mitogenomics
Type of data	Mitogenome sequence data in FASTA file format, tables, mitogenome map in figure format (.PNG), phylogenetic tree in figure format (.PNG) and newick format (.nwk)
How data were acquired	BGISEQ-500RS High-throughput sequencing kit (PN: 85-05238-01, BGI)
Data format	Raw and analyzed
Parameters for data collection	A small amount of tissue from the tip of the tail of Hippocampus kuda complex was sampled, genomic DNA was extracted using Qiagen Blood and Tissue Kit (Qiagen, Valencia, CA), hardware used for quality check includes Qubit 2.0 Fluorometer and Agilent 2100 Bioanalyzer, library preparation prior to sequencing required DNA fragmentation using M220 Focused-ultrasonicator (Covaris, USA), the sample was sequenced using BGISEQ-500RS High-throughput sequencing kit (PN: 85-05238-01, BGI).
Description of data collection	The complete mitogenomes were assembled by using MITOBIM v1.8. The mitogenome mapping quality was assessed using PALEOMIX. The mitogenomes were annotated using MitoAnnotator and GB2sequin annotation web application. The circular mitochondrial genome map was drawn using OGDRAW. Phylogenetic relationship between <i>Hippocampus</i> sp. was constructed using MEGAX.
Data source location	These individuals were caught as incidental catch in fisherman nets at Pulau River, Johor, Malaysia (Latitude: 1° 22' 59.99" N Longitude: 103° 31' 59.99" E)
Data accessibility	The mitogenome data is available in Genbank with the accession numbers MT221436.1 (https://www.ncbi.nlm.nih.gov/nucleotide/MT221436.1) and MT221437.1 (https://www.ncbi.nlm.nih.gov/nucleotide/MT221437.1) and Mendeley data (http://dx.doi.org/10.17632/b3yjvcn7k2.1) [1].
Related research article	S.A. Lourie, R.A. Pollom, S.J. Foster, A global revision of the Seahorses Hippocampus Rafinesque 1810 (Actinopterygii: Syngnathiformes): Taxonomy and biogeography with recommendations for further research, Zootaxa. 4146 (2016) 1–66.

Value of the Data

- The mitogenomes will be useful for *H. kuda* species monitoring using water environmental DNA approach
- The data generated will be useful to resolve the *H. kuda* complex phylogenetic, population and evolutionary studies.
- The data will contribute to our understanding of any adaptive introgression which take place within the *H. kuda* clade.

1. Data Description

The spotted seahorse, *Hippocampus kuda* Bleeker, 1852a is known for its species-complex due to the exceptionally large distribution all around the world [2]. However, among the *Hippocampus* genus, this species is decreasing due to overexploitation for its alleged medicinal properties [3]. Globally, seahorse populations are suffering an exponential decline due to anthropogenic and environmental actions that threaten their survival [3]. Massive development of coastal areas in Malaysia for mega urbanization projects, in regions that serve as its natural habitat are also clearly a threat for its populations [4]. Currently, this species is listed as vulnerable under the (IUCN) Red List of Threatened Species [5].

Here, we provide the Malaysian *H. kuda* mitogenomes with 16,529bp in length. The data information for each individual is presented in Table 1. The representative complete mitogenome map in Fig. 1 shows similar gene arrangement containing 37 genes; 13 protein-coding genes (PCGs), 22 tRNA genes, two rRNA genes, and a non-coding A+T rich control region (D-loop) as in other seahorse mitogenomes [6]. Total length of 13 PCGs is 11,319 bp and they encoded 3773 amino acids. The overall base composition is estimated to be 32.46% for A, 29.40% for T, 14.73% for G and 23.41% for C, indicating an obvious AT rich feature (61.86%). The genes of NAD6 and eight tRNAs are encoded on the light strand, while the rest of mitochondrial genes are encoded on H-strand (Table 2, supplementary data 1).

A phylogenetic tree of all available *Hippocampus* mitogenomes was also constructed (Fig. 2). In total we included the eighteen *Hippocampus* species available in Genbank along with both Malaysian *H. kuda* generated in this work. The mitogenomes were firstly aligned using MUSCLE [7], after which a phylogenetic tree was constructed using the neighbor-joining (NJ) method. The 18 mitogenomes include; *Hippocampus kuda* (accession no. NC_010272.1), *Hippocampus comes* (accession no. NC_020336.1), *Hippocampus trimaculatus* (accession no. NC_021107.1), *Hippocampus histrix* (accession no. NC_021454.1), *Hippocampus erectus* (accession no. NC_022722.1), *Hippocampus ingens* (accession no. NC_024530.1), *Hippocampus barbouri* (accession no. NC_024536.1), *Hippocampus reidi* (accession no. NC_027931.1), *Hippocampus abdominalis* (accession no. NC_028181.1), *Hippocampus kelloggi* (accession no. NC_029349.1), *Hippocampus spinosissimus* (accession no. NC_029350.1), *Hippocampus mohnikei* (accession no. NC_030251.1), *Hippocampus queenslandicus* (accession no. NC_034319.1), *Hippocampus sindonis* (accession no. NC_035827.1), *Hippocampus jayakari* (accession no. NC_036049.1), *Hippocampus camelopardalis* (accession no. NC_041429.1), *Hippocampus capensis* (accession no. NC_042791.1), *Hippocampus hippocampus* (accession no. NC_045033.1). A mitogenome of a pipefish, *Solenostomus paradoxus* (accession no. NC_024186.1) was selected as an outgroup. The phylogenetic tree indicates that the Malaysian *H. kuda* (HK1 and HK2) firstly clustered with *H. capensis*, followed by *H. reidi* and *H. kuda*.

We also compared the mitogenomes to Genbank using BLASTn, and found the closest match for both Malaysian *H. kuda* mitogenomes was a 99.18% similarity to a *H. capensis* (NC_042791.1) sample collected from Bozhou Chinese herbal medicine market (Bozhou, China) [8]. The next closest match, at 97.66%, was to the sole *H. kuda* mitogenome currently available in Genbank (NC_010272.1), from a sample originating Vancouver Aquarium, Canada [9]. A similar match was also found to a *H. reidi* sample (NC_027931.1) [10]. Interestingly, both *H. capensis* and *H. reidi* species are not found in the Malaysian region. However, it is worth noting that there is an

Table 1
Sequencing data for each *H. kuda* individual

	<i>Hippocampus kuda</i> (HK1)	<i>Hippocampus kuda</i> (HK2)
Trimmed reads	119,174,870	34,830,940
Mapped reads	17293	25180
% Mapped reads	0.01	0.07
Depth of coverage (x)	63.19	84.43

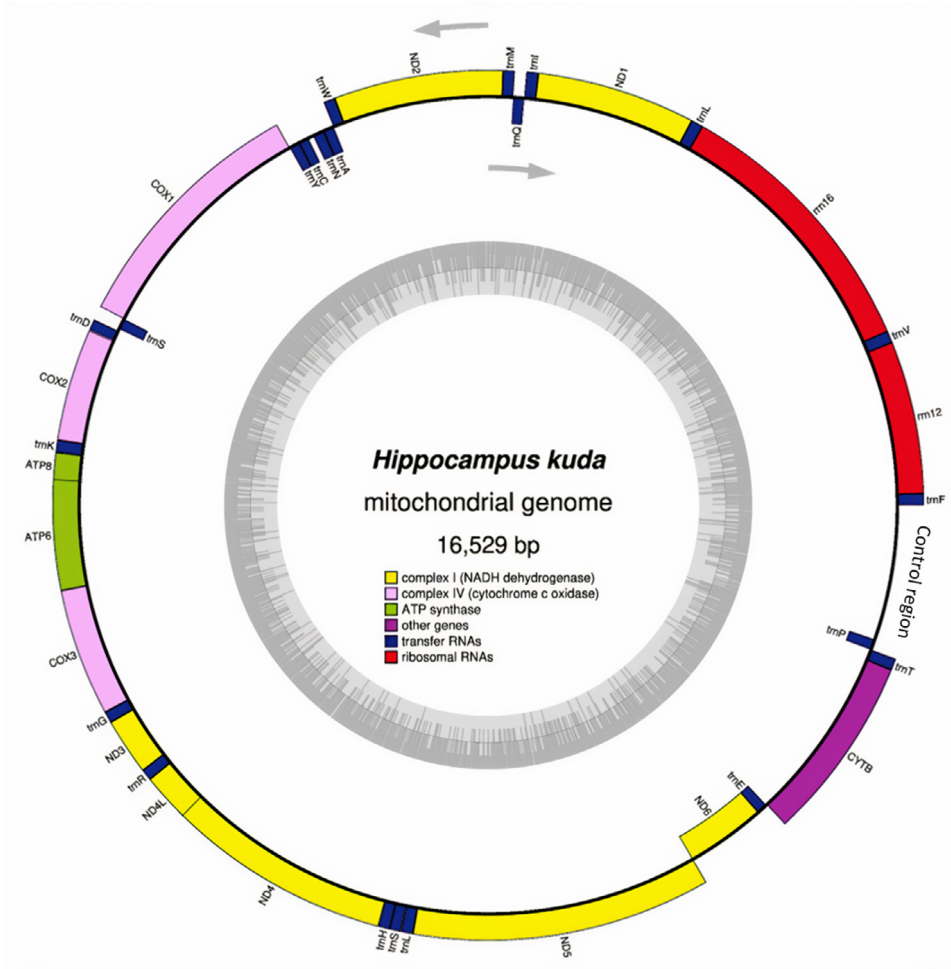


Fig. 1. Map of the *Hippocampus kuda* mitochondrial genome. Genes encoded by the heavy strand shown outside the circle, and those encoded by the light strand are shown inside. The direction of the arrows shows the genes inside the circle are transcribed clockwise and genes outside the circle transcribed counter clockwise. The inner ring shadow indicates the GC content of the genome.

going debate about these species being associated with the unresolved ‘*H. kuda* clade’. Due to their large distribution, these species exhibit localized haplotypes, phylogeographic structuring, and variable morphology [2]. These findings clearly underscore for future studies using nuclear DNA (nuDNA) to fully resolve the relationship within these *Hippocampus* species.

2. Experimental Design, Materials, and Methods

2.1. Biological samples

Two individuals of *Hippocampus kuda* were caught incidentally as bycatch at Pulai River, Johor, Malaysia (Latitude: 1° 22' 59.99" N Longitude: 103° 31' 59.99" E), and identified based on its morphometric features [2]. Tissue samples of the two individuals, *H. kuda* (HK1) and *H. kuda*

Table 2Features of the mitochondrial genome of *Hippocampus kuda*

Gene	Position		Size (bp)	Amino acid	Strand
	From	To			
<i>tRNA(Phe)</i>	1	71	71		H
<i>12S rRNA</i>	72	1009	938		H
<i>tRNA (Val)</i>	1010	1082	73		H
<i>16S rRNA</i>	1083	2778	1696		H
<i>tRNA (Leu)</i>	2779	2852	74		H
<i>NAD1</i>	2853	3827	975	325	H
<i>tRNA (Ile)</i>	3829	3900	72		H
<i>tRNA (Gln)</i>	3900	3970	71		L
<i>tRNA (Met)</i>	3972	4041	70		H
<i>NAD2</i>	4042	5080	1039	346	H
<i>tRNA (Trp)</i>	5081	5151	71		H
<i>tRNA (Ala)</i>	5153	5221	69		L
<i>tRNA (Asn)</i>	5223	5295	73		L
<i>tRNA (Cys)</i>	5331	5396	66		L
<i>tRNA (Tyr)</i>	5397	5463	67		L
<i>COXI</i>	5465	7018	1554	518	H
<i>tRNA (Ser)</i>	7020	7090	71		L
<i>tRNA (Asp)</i>	7105	7172	68		H
<i>COX2</i>	7177	7867	691	230	H
<i>tRNA (Lys)</i>	7868	7942	75		H
<i>ATP8</i>	7944	8111	168	56	H
<i>ATP6</i>	8102	8784	683	228	H
<i>COX3</i>	8785	9568	784	261	H
<i>tRNA (Gly)</i>	9569	9638	70		H
<i>NAD3</i>	9639	9987	349	116	H
<i>tRNA (Arg)</i>	9988	10056	69		H
<i>NAD4-L</i>	10057	10353	297	99	H
<i>NAD4</i>	10347	11725	1379	460	H
<i>tRNA (His)</i>	11728	11796	69		H
<i>tRNA (Ser)</i>	11797	11864	68		H
<i>tRNA (Leu)</i>	11867	11939	73		H
<i>NAD5</i>	11940	13775	1836	612	H
<i>NAD6</i>	13772	14293	522	174	L
<i>tRNA (Glu)</i>	14294	14362	69		L
<i>COB</i>	14367	15507	1141	380	H
<i>tRNA (Thr)</i>	15508	15579	72		H
<i>tRNA (Pro)</i>	15579	15648	70		L
Control region	15648	16529	882		

(HK2) were collected from the tip of the tails. The genomic DNA was extracted using Qiagen Blood and Tissue Kit (Qiagen, Valencia, CA). The DNA was later fragmented into 300–400bp using a M220 Focused-ultrasonicator (Covaris, USA) [11] and BGISEq compatible shotgun sequencing libraries were build using the Blunt-End-Single-Tube (BEST) library protocol [12]. Quantitative PCR was performed prior to index PCR in order to ensure the library was not over-amplified prior to sequencing. Each library was purified using Solid-Phase Reversible Immobilization (SPRI) bead solution. The quality control of generated libraries was quantified quantitatively and qualitatively using Qubit 2.0 Fluorometer (Invitrogen, Merelbeke, Belgium) and Agilent 2100 Bioanalyzer (Agilent, Santa Clara, USA). The libraries were pooled to the equimolar with 15 other libraries (not related to this work). Next, the libraries were sent for shotgun sequencing on the BGISEQ-500 platform in 100bp paired-end mode (PE100) (BGI, Shenzhen, China). The data generated were firstly demultiplexed by index prior to mitogenome construction.

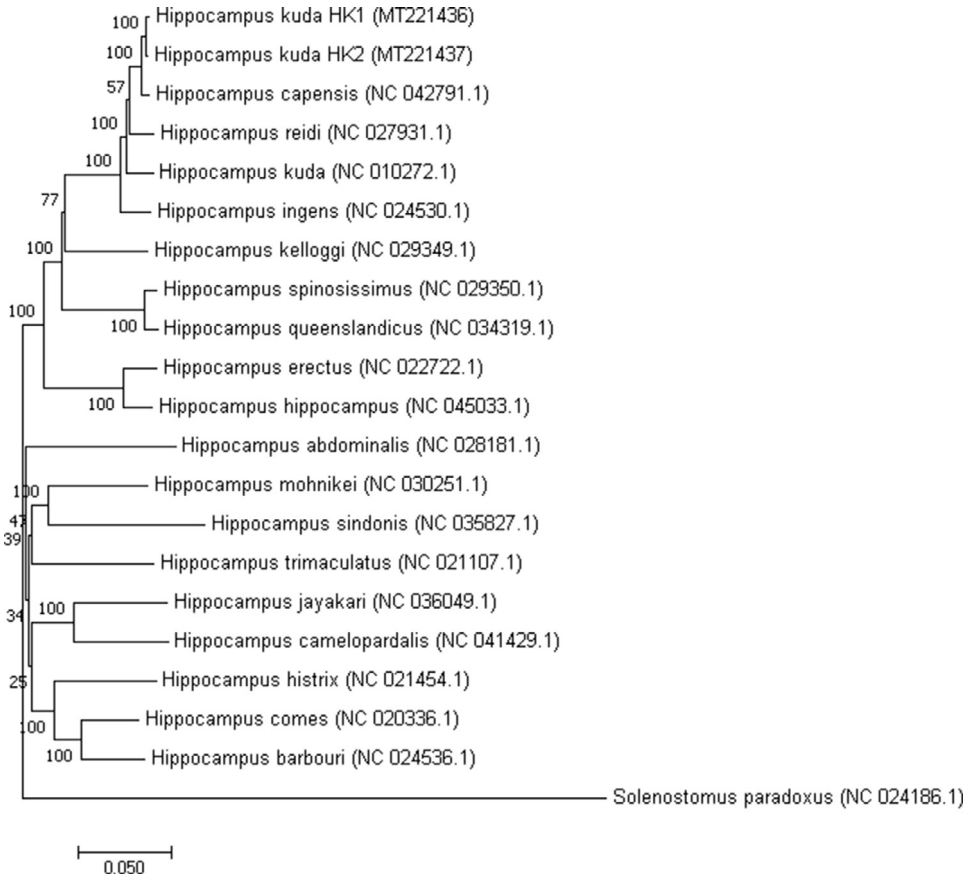


Fig. 2. Phylogenetic tree of two Malaysian *H. kuda* (MT221436 and MT221437) and 18 *Hippocampus* genus constructed with the combined protein-coding gene nucleotide sequences using MEGAX [24]. The tree was generated from NJ method using pipefish as an outgroup. Bootstrap values generated from 1000 replicates for NJ analysis. The number at each node indicated the bootstrap probability of NJ analysis.

2.2. Complete mitogenome generation

The quality of the raw reads generated was verified using the fastQC program (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). The raw reads were trimmed for sequencing adapters, low-quality stretches, and leading/trailing Ns using AdapterRemoval v2.2.2 [13]. Forward and reverse reads were interleaved into a single file prior to the assembly. The assembly *H. kuda* (HK1) and *H. kuda* (HK2) was conducted using MITOBIM v1.8 [14] (default k-mer size of 31), which performs reference assemblies using MIRA iterations [15]. The reference sequence used for the assembly was *H. kuda* from Vancouver Aquarium, Canada (Genbank Accession Number: NC_010272.1). Next, we used the PALEOMIX v1.2.6 BAM pipeline [16] with default parameters to remove reads shorter than 25 bp after trimming. The trimmed reads were aligned using Burrows-Wheeler Aligner [17] against the newly assembled mitogenome constructed by MITOBIM. Further trimming for the alignments that showed PCR duplicates and low-quality scores were conducted using MarkDuplicates program from Picard tools [18]. Next, the IndelRealigner tool from the Genome Analysis Toolkit (GATK) [19] was used to locally realign the reads around the small insertions and deletions (indels) in order to improve overall genome quality. Post-analysis, the statistics of the sequencing data for each individual was generated as displayed in

Table 1. Tablet software [20] was used to manually check the indels and read coverage along the assembled mitogenomes. The mitogenome was annotated using the MitoAnnotator [21] and GB2sequin annotation web application [22]. The circular mitochondrial genome map was drawn using OGDRAW [23] (Fig. 1).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

Acknowledgments

The authors would like to thank Stine R. Richter for her assistance during library preparation and Ashot Margaryan for guidance on bioinformatics analysis.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi: [10.1016/j.dib.2020.105721](https://doi.org/10.1016/j.dib.2020.105721).

Appendix A. Supplementary data 1

Features of the mitochondrial genome of *Hippocampus kuda*

References

- [1] [dataset] PUTERI NUR SYAHZANANI AHARI, Abdul Malik, Nur Fatimah, Mohd Shahir Shamsir Omar, M Gilbert, P. Thomas, Faezah Mohd Salleh, "The First Complete Mitochondrial Genome Data of Hippocampus kuda Originating from Malaysia", Mendeley Data (2020). v1,2020 <http://dx.doi.org/10.17632/b3yjvcn7k2.1>.
- [2] S.A. Lourie, R.A. Pollom, S.J. Foster, A global revision of the Seahorses Hippocampus Rafinesque 1810 (Actinopterygii: Syngnathiformes): Taxonomy and biogeography with recommendations for further research, *Zootaxa* 4146 (2016) 1–66.
- [3] M.L.F. Ternes, L.C. Gerhardinger, A. Schiavetti, Seahorses in focus: local ecological knowledge of seahorse-watching operators in a tropical estuary, *J. Ethnobiol. Ethnomed.* 12 (2016) 52.
- [4] J.M. Salim, M.A. Radzi, S.M. Razali, F.M. Cooke, Coastal Landscapes of Peninsular Malaysia: The Changes and Implications for Their Resilience and Ecosystem Services, in: *Landscape Reclamation-Rising From What's Left*, IntechOpen (2018).
- [5] IUCN, Hippocampus kuda: Aylesworth, L.: The IUCN Red List of Threatened Species 2014: e.T10075A16664386, (2012). <https://doi.org/10.2305/IUCN.UK.2014-3.RLTS.T10075A16664386.en>.
- [6] M. Lai, S. Sun, J. Chen, Z. Lou, F. Qiu, G. Zhang, R. Cheng, Complete mitochondrial genome sequence for the seahorse Adulteration Hippocampus camelopardalis Bianconi 1854, *Mitochondrial DNA Part B* 4 (2019) 432–433.
- [7] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res* 32 (2004) 1792–1797.
- [8] Y. Ge, L. Zhu, M. Chen, G. Zhang, Z. Huang, R. Cheng, Complete mitochondrial genome sequence for the endangered Knysna seahorse Hippocampus capensis Boulenger 1900, *Conserv. Genet. Resour.* 10 (2018) 461–465.
- [9] R. Kawahara, M. Miya, K. Mabuchi, S. Lavoué, J.G. Inoue, T.P. Satoh, A. Kawaguchi, M. Nishida, Interrelationships of the 11 gasterosteiform families (sticklebacks, pipefishes, and their relatives): a new perspective based on whole mitogenome sequences from 75 higher teleosts, *Mol. Phylogenet. Evol* 46 (2008) 224–236.
- [10] X. Wang, Y. Zhang, H. Zhang, T. Meng, Q. Lin, Complete mitochondrial genome sequence of the longsnout seahorse Hippocampus reidi (Ginsburg, 1933; Gasterosteiformes: Syngnathidae), *Mitochondrial DNA A DNA Mapp Seq Anal* 27 (2016) 1401–1402.
- [11] F. Mohd Salleh, J. Ramos-Madriral, F. Peñaloza, S. Liu, S.S. Mikkel-Holger, P.P. Riddhi, R. Martins, D. Lenz, J. Fickel, C. Roos, M.S. Shamsir, M.S. Azman, K.L. Burton, J.R. Stephen, A. Wiltling, M.T.P. Gilbert, An expanded mammal mitogenome dataset from Southeast Asia, *Gigascience* 6 (2017) 1–8.
- [12] C. Carøe, S. Gopalakrishnan, L. Vinner, S.S.T. Mak, M.H.S. Sinding, J.A. Samaniego, N. Wales, T. Sicheritz-Pontén, M.T.P. Gilbert, Single-tube library preparation for degraded DNA, *Methods Ecol. Evol* 9 (2018) 410–419.

- [13] M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: rapid adapter trimming, identification, and read merging, *BMC Res. Notes* 9 (2016) 88.
- [14] C. Hahn, L. Bachmann, B. Chevreux, Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach, *Nucleic Acids Res* 41 (2013) e129–e129.
- [15] B. Chevreux, MIRA: an automated genome and EST assembler, (2007). https://archiv.ub.uni-heidelberg.de/volltextserver/7871/1/thesis_zusammenfassung.pdf.
- [16] M. Schubert, L. Ermini, C. Der Sarkissian, H. Jónsson, A. Ginolhac, R. Schaefer, M.D. Martin, R. Fernández, M. Kircher, M. McCue, E. Willerslev, L. Orlando, Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX, *Nat. Protoc.* 9 (2014) 1056–1082.
- [17] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform, *Bioinformatics* 25 (2009) 1754–1760.
- [18] A. Wysoker, K. Tibbetts, T. Fennell, Picard tools version 1.90, [Http://picard. Sourceforge. Net](http://picard.sourceforge.net) (Accessed 14 December 2016). 107 (2013) 308.
- [19] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernysky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M.A. DePristo, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data, *Genome Res* 20 (2010) 1297–1303.
- [20] I. Milne, M. Bayer, L. Cardle, P. Shaw, G. Stephen, F. Wright, D. Marshall, Tablet—next generation sequence assembly visualization, *Bioinformatics* 26 (2010) 401–402.
- [21] W. Iwasaki, T. Fukunaga, R. Isagozawa, K. Yamada, Y. Maeda, T.P. Satoh, T. Sado, K. Mabuchi, H. Takeshima, M. Miya, M. Nishida, MitoFish and MitoAnnotator: a mitochondrial genome database of fish with an accurate and automatic annotation pipeline, *Mol. Biol. Evol.* 30 (2013) 2531–2540.
- [22] P. Lehwark, S. Greiner, GB2sequin - A file converter preparing custom GenBank files for database submission, *Genomics* 111 (2019) 759–761.
- [23] M. Lohse, O. Drechsel, S. Kahlau, R. Bock, OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets, *Nucleic Acids Res* 41 (2013) W575–W581.
- [24] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms, *Mol. Biol. Evol.* 35 (2018) 1547–1549.