

## The near-complete mitogenome of the critically endangered *Pseudocleopatra dartavellei* (Caenogastropoda: Paludomidae) from the Congo River assembled from historical museum material

Björn Stelbrink<sup>a</sup> , Christian Kehlmaier<sup>b</sup> , Thomas Wilke<sup>a</sup>  and Christian Albrecht<sup>a</sup> 

<sup>a</sup>Department of Animal Ecology and Systematics, Justus Liebig University Giessen, Giessen, Germany; <sup>b</sup>Museum of Zoology, Senckenberg Dresden, Dresden, Germany

### ABSTRACT

Here, we present the first near-complete mitogenome of a member of the freshwater gastropod family Paludomidae, *Pseudocleopatra dartavellei*. This Congo River species is of particular importance because the sister to the Lake Tanganyika radiation is supposed to be a paludomid riverine species. We used ancient DNA (aDNA) techniques including single-stranded DNA library preparation in order to assemble the mitogenome from historical museum material collected in 1937. The mitogenome was 15,368 bp long and showed typical characteristics as identified in other freshwater gastropods. The present phylogeny shows a closer relationship between *Pseudocleopatra dartavellei* and another non-Tanganyikan species, *Cleopatra johnstoni*.

### ARTICLE HISTORY

Received 24 June 2019  
Accepted 16 August 2019

### KEYWORDS

Africa; Democratic Republic of the Congo; freshwater gastropods; ancient DNA; museum genomics

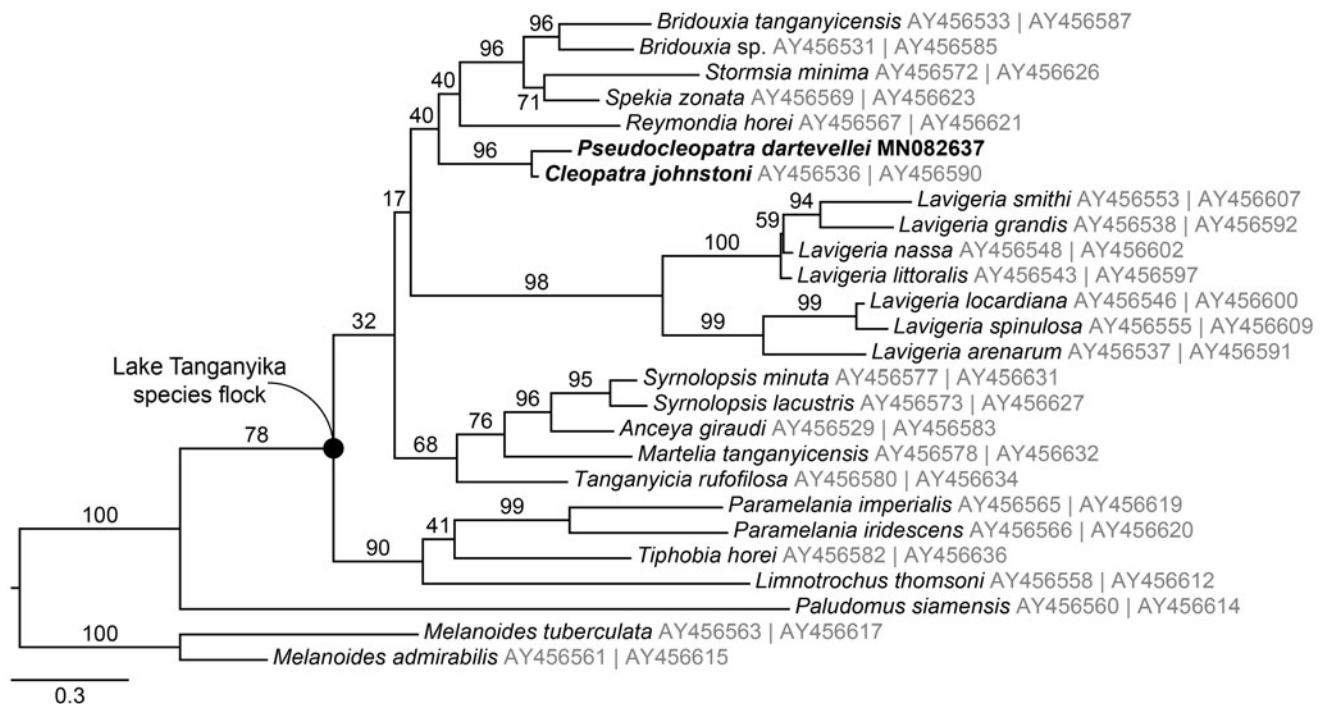
Lake Tanganyika is home to a famous radiation of freshwater gastropods of the family Paludomidae (17 genera; Brown 1994; Neiber and Glaubrecht 2019). There is an ongoing discussion on whether Lake Tanganyika may have acted as a refuge for ancient lineages that existed in pre-lake systems such as the Congo River (Wilson et al. 2004). However, studies aiming at uncovering the biogeographical origin of the species flock have been hampered by the inaccessibility of key areas in the Congo Basin due to civil wars and unrest (Glaubrecht and Strong 2007). Under such circumstances, material conserved in institutional collections such as museums proves very valuable. This is also the case for the critically endangered riverine genus *Pseudocleopatra* (see Jørgensen 2010), which occurs in western Africa and parts of the Congo basin and which represents one of the potential sister groups of the Lake Tanganyika gastropod radiation.

A specimen of *Pseudocleopatra dartavellei* Mandahl-Barth, 1973 was collected by Edmond Dartavelle in the Congo River near Matadi, Democratic Republic of the Congo (formerly Belgian Congo; 5.8167°S 13.4667°E) in 1937 and is deposited in the Royal Museum of Central Africa, Tervuren, Belgium (RMCA 233464). The material was analyzed applying NGS protocols developed for ancient and heavily degraded DNA in the cleanroom facility of the Senckenberg Natural History Collections, Dresden.

Genomic DNA was extracted from c. 3 mm<sup>3</sup> of soft tissue using the GEN-IAL All-tissue DNA-Kit (GEN-IAL GmbH, Troisdorf, Germany) protocol for forensic material and converted into a single-indexed single-stranded Illumina

sequencing library (Gansauge and Meyer 2013; Korlević et al. 2015), including an uracil-DNA glycosylase (UDG) treatment. Shotgun sequencing was conducted on an Illumina MiSeq<sup>®</sup> platform (75 bp paired-end reads). After adapter trimming, quality filtering, and duplicate removal, reads mapping to a set of non-mollusk mitogenomes were excluded using FastQ Screen 0.11.4 (Wingett et al. 2018). The reduced readpool (c. 13.4 million reads) was then mapped against *Cerithidea obtusa* (NC\_039951) using BWA 0.7.15 (Li and Durbin 2009). The obtained fragmentary consensus was used as a seed reference for a two-step baiting and iterative mapping approach in MITObim 1.9 (Hahn et al. 2013), with an allowed mismatch value of 2, resulting in a single high-quality contig (assembled reads: 24,308; average read length: 51 bp; maximum coverage = 211; average coverage = 82). The contig was visualized and checked for coverage and assembly artifacts in Tablet 1.16.09.06 (Milne et al. 2013), and the authenticity of the mapped reads was tested with mapDamage 2.0 (Jónsson et al. 2013). The contig was finally annotated on the MITOS2 Web Server (Bernt et al. 2013; reference dataset = RefSeq 63 Metazoa).

The almost complete, annotated mitogenome has 15,368 bp (with only three ambiguous sites) and is deposited at GenBank (MN082637). The base composition was biased towards a high A + T content (63.72%; A = 28.16%, T = 35.56%), whereas the G + C content was comparatively low (36.26%; G = 19.44%, C = 16.82%). The majority of protein-coding genes (PCGs) started with an ATG (i.e. AUG) codon and stopped with a TAA codon. Gene arrangement on



**Figure 1.** COX1–16S-rRNA maximum-likelihood phylogeny based on GenBank data (Wilson et al. 2004). The non-Tanganyikan species *Pseudocleopatra dartevellei* (Congo River) and *Cleopatra johnstoni* (Lake Mweru, Zambia) are marked in bold; the two *Melanoides* species were used as outgroups. GenBank accession numbers are shown in grey (left: COX1, right: 16S-rRNA).

the light strand was: *tRNA<sup>Cys</sup>*, *tRNA<sup>Arg</sup>*, *tRNA<sup>Ala</sup>*, *tRNA<sup>Asn</sup>*, *tRNA<sup>Trp</sup>*, *tRNA<sup>Glu</sup>*, *tRNA<sup>Tyr</sup>*, *tRNA<sup>Lys</sup>*, *COX3*, *tRNA<sup>Met</sup>*, *COB*, *ND6*, *tRNA<sup>Pro</sup>*, *ND1*, *tRNA<sup>Leu2</sup>*, *tRNA<sup>Leu1</sup>*, *16S-rRNA*, *tRNA<sup>Val</sup>*, *tRNA<sup>Gly</sup>*, *tRNA<sup>Thr</sup>*, and *12S-rRNA*. On the heavy strand, gene order was: *tRNA<sup>Ser1</sup>*, *ND2*, *tRNA<sup>Asp</sup>*, *ATP8*, *ATP6*, *tRNA<sup>Ile</sup>*, *ND3*, *COX1*, *COX2*, *tRNA<sup>Ser2</sup>*, *tRNA<sup>Gln</sup>*, *ND4L*, *ND4*, *tRNA<sup>His</sup>*, *ND5*, and *tRNA<sup>Phe</sup>*. Overlaps were identified between *COB* and *ND6* (47 bp), *tRNA<sup>Thr</sup>* and *12S-rRNA* (3 bp), *ND2* and *tRNA<sup>Asp</sup>* (2 bp), and *ND5* and *tRNA<sup>Phe</sup>* (1 bp).

For a phylogenetic reconstruction of the Paludomidae, available mitochondrial DNA sequences were downloaded from GenBank that refers to the study of Wilson et al. (2004). A maximum-likelihood tree based on partial sequences of *COX1* and *16S-rRNA* was constructed (Figure 1) using RAXML-HPC BLACKBOX 8.2.10 (Stamatakis 2014; settings: GTR +  $\Gamma$  model, codon partition scheme) on the CIPRES Science Gateway (Miller et al. 2010). Accordingly, *Pseudocleopatra dartevellei* is sister to *Cleopatra johnstoni* from Lake Mweru, a lake in the southeastern corner of the Congo drainage system. Although intergeneric relationships are often poorly supported, the present phylogeny supports a previous study that suggested a non-monophyly of the Lake Tanganyika paludomids (see Wilson et al. 2004).

## Disclosure statement

The authors report no conflict of interest.

## Acknowledgements

We are especially grateful to Christophe Allard (Royal Museum for Central Africa, Tervuren, Belgium) for providing material of *Pseudocleopatra dartevellei* and other African species. Laboratory work

was conducted at the cleanroom facility of the Museum of Zoology, Senckenberg Dresden (SGN-SNSD-Mol-Lab).

## Funding

BS was supported by a Deutsche Forschungsgemeinschaft grant [DFG STE 2460/2-1] on diversification dynamics in ancient lake gastropod species flocks.

## ORCID

Björn Stelbrink  <http://orcid.org/0000-0002-7471-4992>  
 Christian Kehlmaier  <http://orcid.org/0000-0001-9622-0566>  
 Thomas Wilke  <http://orcid.org/0000-0001-8263-7758>  
 Christian Albrecht  <http://orcid.org/0000-0002-1490-1825>

## References

- Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritsch G, Pütz J, Middendorf M, Stadler PF. 2013. MITOS: improved *de novo* metazoan mitochondrial genome annotation. *Mol Phylogenet Evol.* 69:313–319.
- Brown DS. 1994. *Freshwater snails of Africa and their medical importance*. London (UK): Taylor and Francis.
- Gansauge MT, Meyer M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat Protoc.* 8: 737–748.
- Glaubrecht M, Strong EE. 2007. Ancestry to an endemic radiation in Lake Tanganyika? Evolution of the viviparous gastropod *Potadomoides* Leloup, 1953 in the Congo River system (Caenogastropoda, Cerithioidea, Paludomidae). *Biol J Linn Soc.* 92:367–401.
- Hahn C, Bachmann L, Chevreux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Res.* 41:e129.
- Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. 2013. MapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics.* 29:1682–1684.

- Jørgensen A. 2010. *Pseudocleopatra dartevellei*. The IUCN Red List of Threatened Species 2010: e.T184587A8297083. <http://dx.doi.org/10.2305/IUCN.UK.2010-3.RLTS.T184587A8297083.en>. Downloaded on 28 May 2019.
- Korlević P, Gerber T, Gansauge M, Hajdinjak M, Nagel S, Aximu-Petri A, Meyer M. 2015. Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques*. 59:87–93.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 25:1754–1760.
- Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop; Nov 14; Piscataway (NJ): IEEE.
- Milne I, Stephen G, Bayer M, Cock PJA, Pritchard L, Cardle L, Shaw PD, Marshall D. 2013. Using tablet for visual exploration of second-generation sequencing data. *Brief Bioinform*. 14:193–202.
- Neiber MT, Glaubrecht M. 2019. Annotated nomenclator of extant and fossil taxa of the Paludomidae (Caenogastropoda, Cerithioidea). *Zookeys*. 850:1–132.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30:1312–1313.
- Wilson AB, Glaubrecht M, Meyer A. 2004. Ancient lakes as evolutionary reservoirs: evidence from the thalassoid gastropods of Lake Tanganyika. *Proc R Soc Lond B*. 271:529–536.
- Wingett SW, Andrews S, Hamilton RS. 2018. FastQ Screen: a tool for multi-genome mapping and quality control. *F1000Res*. 7:1338.