Mitogenomic characterisation and systematic placement of the Congo blind barb *Caecobarbus geertsii* (Cypriniformes: Cyprinidae)

Nele Mullens[1*], Gontran Sonet[2], Eva Decru[1,3], Massimiliano Virgilio[1], Snoeks Jos[1,3], Emmanuel Vreven[1,3*]

[1] Royal Museum for Central Africa, Biology Department, Leuvensesteenweg 13, B-3080 Tervuren, Belgium

[2] Royal Belgian Institute of Natural Sciences, JEMU, Vautierstraat 29, B-1000 Brussels, Belgium

[3] KU Leuven, Laboratory of Biodiversity and Evolutionary Genomics, Charles Deberiotstraat 32, B-3000 Leuven, Belgium

Gontran Sonet: gsonet@naturalsciences.be

Eva Decru: eva.decru.icht@gmail.com

Massimiliano Virgilio: massimiliano.virgilio@africamuseum.be

Snoeks Jos: jos.snoeks@africamuseum.be

*corresponding authors

Emmanuel Vreven: emmanuel.vreven@africamuseum.be

Nele Mullens: nele.mullens@africamuseum.be

Abstract

This study presents the first complete mitochondrial genome (mitogenome) of *Caecobarbus geertsii*, the Congo blind barb, a cave-dwelling, CITES-protected, cyprinid fish endemic to the Lower Congo basin (DRC). The length of the circular mitogenome is 16,565 base pairs. The 13 protein coding genes, 2 ribosomal RNA genes and 22 transfer RNA genes are similar in position and direction to those of other members of the family Cyprinidae. Phylogenetic analyses including 28 complete mitogenomes from representatives of the subfamily Smiliogastrinae (Cyprinidae), showed that *Caecobarbus* was nested within a clade including representatives of the genus *Enteromius.* The data presented in this study provide information on the molecular identification and classification of this threatened species. The results further suggest the need for a taxonomic revision of the genus *Enteromius*.

Keywords: Smiliogastrinae, mitogenomics, molecular taxonomy

1. Introduction

Caecobarbus geertsii Boulenger, 1921 [1], also known as the Congo blind barb or 'Nzonzi a mpofo' in the local Kikongo language, is a cave-dwelling cyprinid fish. It is only found in the Lower Congo basin [2] and belongs to the monotypic genus *Caecobarbus* Boulenger 1921. This genus is placed within the species-rich family of the Cyprinidae [3,4], which currently includes 159 genera and over 1744 valid species [4,5]. Currently, already more than 145 complete mitogenomes are available for the Cyprinidae on online repositories such as the National Center for Biotechnology Information (NCBI) and the Barcode Of Life Database (BOLD). *Caecobarbus* is currently assigned to Cyprinidae as *incertae sedis* by Tan and Armbruster [3].

The species is a true troglobiont, living permanently in caves [6]. Its known distribution is currently limited to seven caves in the Mbanza-Ngungu region in the Lower Congo River basin in the Democratic Republic of the Congo (DRC) [2]. Of all teleostean fishes, only 12 families have representatives that are adapted to living in the extreme conditions of caves, with little or no light and, generally, reduced food availability. Cyprinidae typically have a diurnal activity, suggesting that a number of morphological and life history traits of *C. geertsii* such as the loss of functional eyes and pigmentation, and limited growth, represent adaptive responses of troglobionts to cave conditions [2,7,8]. *Caecobarbus geertsii*, which in the past was traded as an ornamental aquarium fish [2] is listed as a CITES II species [9] (Convention on International Trade in Endangered Species of Wild Fauna and Flora, trade is restricted to prevent risk of extinction). It is currently threatened by anthropogenic disturbances such as irrigation and deforestation, which lead to changes in hydrogeological conditions including those of the underground rivers feeding the caves, and which can modify siltation patterns within these caves [2,10].

In the past, most of the African barbs were attributed to the genus *Barbus*, which has been

described as a “monstrous aggregation” [11]. However, karyological research undertaken in the 1990’s, revealed the genus to be an aggregate of species with three different levels of ploidy [12–14]: (i) diploids, containing the small African ‘*Barbus*’; (ii) tetraploids, including the type species, and thus identified as *Barbus sensu stricto*, and (iii) hexaploids, now referred to as *Labeobarbus* with its junior synonym *Varicorhinus* [15–17]. Karyological research also demonstrated *C. geertsii* to be diploid (2n=50) [18]. More recently, phylogenetic reconstructions based on mitochondrial DNA [3] revealed that, while European tetraploid species identified as *Barbus sensu stricto* are part of the subfamily Barbinae, the African diploid and hexaploid species are part of the subfamilies Smiliogastrinae and Torinae respectively [3]. This lead to the revalidation of *Enteromius* Cope, 1867 [19], the oldest available name for the African diploid species [11,12,20]. Being diploid [18] we expect *Caecobarbus* to be part of the subfamily Smiliogastrinae. Currently, besides a relatively limited number of reference DNA barcodes (see below), there is no further genetic information available for *C. geertsii* in public repositories such as the NCBI. This study provides a first molecular characterization of the complete mitogenome of *C. geertsii* and explores its taxonomic placement and phylogenetic relationships within the subfamily Smiliogastrinae.

2. Material and Methods

The specimen (MRAC 2017-015-P-0015) sequenced in this study was collected with a dip net in a cave in the Mbanza-Ngungu region (Grotte de Kiamvu, village Kiasi Mankala (5°49’44.6”S -14°55’57.2”E, WGS84, 15/08/2017, collectors: Vreven E., Kimbembi A., Kosi Zola J. and Wamuini Lunkayilakio S.), situated in Zone I, the northeastern part of the cave system as defined by Heuts and Leleup [21]. After morphological identification, DNA was extracted from a fin clip using the DNeasy Blood &Tissue kit (Qiagen Inc., Hilden,

Germany). DNA (100 ng normalized in 50 µl ultrapure water) was sonicated in a Covaris S220 Focused-ultrasonicator and fragmented to an average size of 350 base pairs (bp). As part of a wider survey, we used the TruSeq Nano DNA LT Library Prep Kit (Illumina) to prepare total genomic libraries including either individually indexed specimens or pooled specimens from different species. In this case, the same index was used for a pooled library including the *C. geertsii* specimen and a Syrphid (*Syritta bulbus,* Diptera, Syrphidae*,* RMCA). Total genomic libraries were sequenced on an Illumina MiSeq platform and the resulting paired end 300 bp reads were processed as described below. The mitochondrial reads of *C. geertsii* and *S. bulbus* (Diptera, Syrphidae) were separated *in silico* via reference mapping using closely related and publicly available mitogenomes. For *C. geertsii* we used the BBMap [22] plugin of Geneious 10.2.3 (https://www.geneious.com) with default parameters, using a reference alignment including the following 15 representatives of the family Cyprinidae for which a full mitogenome was available: *Sinocyclocheilus multipunctatus* (GenBank accession: MG026730), *Gnathopogon nicholsi* (NC_033351), *Garra salweenica* (NC_033389), *Rectoris posehensis* (NC_033390), *Herzensteinia microcephalus* (NC_033403), *Epalzeorhynchos frenatus* (NC_033962), *Rasbora argyrotaenia* (NC_035749), *Xenocypris yunnanensis* (NC_035954), *Rohtee belangeri* (NC_036232), *Discogobio longibarbatus* (NC_036301), *Schizothorax taliensis* (NC_037516), *Microphysogobio kiatingensis* (NC_037402), *Acheilognathus omeiensis* (NC_037404), *Semilabeo obscurus* (NC_037408) and *Candidia barbatus* (NC_037156) (Tab. 1). A first draft annotation of the mitogenome was obtained using the default settings of MITOS [23]. Then, the largest open reading frames (ORFs) found between tRNAs were selected, allowing truncated (incomplete) stop codons (which are completed by RNA polyadenylation) and overlap between adjacent ORFs. The cloverleaf structure of the 22 inferred tRNAs were visualized in MITOS. The assembled and annotated mitogenome was submitted to GenBank with accession number MK457704. Gene symbols

and their concordance with the nomenclature are given in SM 1.

Forty-six publicly available reference DNA barcodes of *C. geerstii* were preliminarily analyzed using the Barcode Index Number (BIN) System [24], i.e. the algorithm for Operational Taxonomic Unit (OTU) delimitation currently implemented on the Barcoding of Life Data System (www.boldsystems.org). The BIN systems aims at providing putative species identification without prior taxonomic information. Phylogenetic analyses were implemented via Bayesian and Maximum likelihood tree reconstructions based on 27 publicly available mitogenomes for the subfamily Smiliogastriae (Tab. 2). Representatives of two closely related subfamilies, *Neolissochilus hexagonolepis* (KM668070) and *Hypselobarbus pulchellus* (KX177967) from the subfamily Torinae, and *Garra orientalis* (JX290078) from the subfamily Labeoninae [3,20] were used as outgroups. As the taxonomic classification of *Barbodes semifasciolatus* is currently a matter of debate [20], we tentatively considered the nomenclature as suggested by Kottelat (2013) [25] based on morphological evidence and currently followed in both Eschmeyer's Catalog of Fishes [26] and FishBase [27]. The 13 protein coding (PCGs) and 2 ribosomal RNA (rRNA) genes of each mitogenome were aligned using the default parameters of ClustalW [28] and concatenated.

PartitionFinder 2.1.1 [29] was used to select the most appropriate substitution model for each gene and ribosomal partition as obtained using linked branch lengths and a greedy search. Model partitioning based on codon position was initially explored and then abandoned as preliminary analyses showed it could not significantly improve node support. Bayesian tree reconstructions, implemented in MrBayes v3 [30], were based on the general time reversible model [31] with an estimated proportion of invariant sites and gamma distributed among-site rate variation (GTR+I+G) for all partitions except atp8, for which PartitionFinder considered the Hasegawa Kishino and Yano substitution model [32] with gamma distributed rates (HKY+G). In the tree reconstruction, MrBayes employed a cold chain and three incrementally

heated chains. Starting trees for each chain were random and the default values of MrBayes were chosen for all settings (including prior distributions). MrBayes metropolis coupled Markov Chains Monte Carlo (MCMC) were run for 50 million generations (until the average standard deviation of split frequencies fell below 0.01) with heating temperature of 0.1. Trees were sampled every 1000 generations with 50% of them discarded as burn-in. Maximum likelihood (ML) analyses were implemented using RAxML-HPC v.8 [33] with autoMRE bootstrapping (setting a maximum of 1000 bootstrap) and GTRGAMMA approximation [33]. A consensus tree was generated from the ML trees and visualized in FigTree v1.4.3 [34]. Nodes with Bayesian posterior probabilities (PP) > 0.95 and bootstrap support > 75% were considered as supported, all unsupported nodes were excluded from further consideration. Bayesian and ML tree reconstructions were obtained using the MrBayes on XSEDE (v3.2.6) and RAxML BlackBox interfaces available on the CIPRES Science Gateway V3.3 portal (www.phylo.org) [35].

3. Results

The full mitogenome of *Caecobarbus geertsii* assembled in this study (SM 1) contains two rRNA genes, 13 protein coding genes (PCGs), 22 tRNA genes and a non-coding AT-rich region (D-loop) (Fig. 1, SM 1), and has an AT bias of 60.4% (Tab. 2). The length of the mitogenome of *C. geertsii* falls in line with that of the 27 other members of the subfamily Smiliogastrinae available (16,565 bp for *Caecobarbus* compared to an average of 16,729 bp (SD=225)) (Tab. 2). All PCGs of *C. geerstii* begin with an ATG start codon, with the exception of COX1 which starts with GTG (SM 1). This is largely in line with what observed in the screening of the other 27 Smiliogastrinae, with the possible exception of NAD3, NAD4L and NAD5 where alternative start codons can be exceptionally observed (Fig. 2A). Only six PCGs in *C. geerstii* (COX1, COX3, ATP6, ATP8, NAD4l and NAD5) terminate with

the typical TAA stop codon, while five other PCGs (NAD2, NAD3, NAD4, COX2, and CytB) terminate with an incomplete codon (T-- or TA-) and the remaining two (NAD1 and NAD6) with TAG (SM 1). Most stop codons are in line with what described for other Smiliogastrinae (Fig. 2B). COX1, COX2, NAD4L and NAD4 share the same stop codon in all species, including *C. geertsii*. For NAD1 the stop codon of *C. geertsii* (TAG) is not in line with what observed in other Smiliogastrinae, where the TAA has been more commonly reported. Additionally, the complete TAA codon observed for COX3 in *C. geertsii* has not been commonly observed in other Smiliogastrinae, which rather have incomplete stop codons (T-- or TA-). The stop codons for NAD6 reported in Smilogastrine are either TAA or TAG (in similar proportions), with the stop codon observed in *C. geertsii* being TAG. In three cases, PCGs on the same strand overlap: atp8-atp6 (seven overlapping bp), atp6-COX3 (one bp) and NAD4l-NAD4 (seven bp). A seven bp overlap between two genes on opposite strands was also observed (NAD5-NAD6). The 12S and 16S rRNA (rrnS and rrnL) are 955 and 1686 bp long. The length of the AT-rich region was 903 bp in length. In total 14 of the 22 tRNA genes, both rRNAs and 12 of the 13 PCGs were located on the H-strand; the remaining eight tRNAs and a single PCG were located on the L-strand (SM 1).

The preliminary screening of the Barcode Index Numbers (BINs) available for the 46 public DNA barcodes of *C. geertsii* revealed that all COX1 sequences of this species cluster together in a single Barcode Index Number (BIN ID = BOLD:AEB2844), whose nearest neighbor (BIN ID = BOLD:ACV4576) includes representatives of genus *Enteromius* (average p-distance within the *Caecobarbus* BIN BOLD:AEB2844 = 0.44, distance to the *Enteromius* BIN BOLD:ACV4576 = 4.49%). The Bayesian and ML tree reconstructions (Fig. 3, SM 2, SM 3) recovered *C. geertsii* within the genus *Enteromius,* in a clade composed of four groups: two sister groups, one including *C. geertsii* and *E. camptacanthus* and the other *E. callipterus, E. eburneensis* and *E. guirali.* The two remaining groups were recovered in basal positions;

one including *E. trimaculatus*, *E. fasciolatus* and *E. pobeguini,* and the other being composed of *E. jae* and *E. hulstaerti*. The genus *Systomus*, represented by *S. orphoides* and *S. sarana*, was recovered as a well-supported sister group to *Caecobarbus*/*Enteromius*.

Phylogenetic relationships between *Caecobarbus*/*Enteromius*/*Systomus* and the other target taxa could only be partially resolved. *Barbodes* is found not to be monophyletic, as *B. semifasciolatus* was recovered in a basal position of the tree as a sister species of *Chagunius chagunio*. The three remaining *Barbodes* representatives (*B. binotatus*, *B. aurotaeniatus, B. lateristriga*) were part of a clade including *Hampala* (*H. macrolepidota*) and *Puntigrus* (*P. tetrazona*), with the latter in a basal position. Even if the genera *Rohtee* (*R. ogibii* and *R. belangeri*), *Sahyadria* (*S. chalakkudiensis* and *S. denisonii*) and *Osteobrama* (*O. faeae, O, cotio, O. cunma*) could be recovered as monophyletic, relationships between these group and the other target taxa (*P. ticto, P. titteya*) remained poorly resolved.

4. Discussion

This study presents the first complete, annotated mitogenome of *Caecobarbus geertsii*, a CITES-protected cyprinid, endemic to the Congo basin. This is furthermore the first complete mitogenome of a Congo basin endemic member of the subfamily Smiliogastrinae. The mitogenome of *C. geertsii* is largely comparable, both in terms of gene order and mitogenome length, with the other mitochondrial genomes described for Smilogastrinae (Tab. 2, SM 1). The mitochondrial phylogeny presented in this study (Fig. 3) allowed recovering *Caecobarbus geertsii* within a highly supported clade including members of the genus *Enteromius*. This confirms the assignment of *Caecobarbus* to the subfamily Smiliogastrinae, as originally hypothesized on the basis of its diploid genome [36] and of its morphological [37,38] and genetic similarity with *Enteromius*. Indeed, more than half a century ago, Heuts [37] further highlighted the morphological similarity between *C. geertsii* and *Enteromius*

(formerly *Barbus*) *holotaenia* Boulenger, 1904, with the former differing from the latter only in the lack of pigmentation and functional eyes.

The morphological distinction of the genus *Caecobarbus* from the former genus *Barbus*, now *Enteromius*, is based on the presumed absence of eyes [1,39] (eyes were later found to be reduced, not absent [2,37]), the lack of striation on scales [1], and, later on, the lack of pigment too [39]. All these characters, however, have been indicated as reductive or regressive [40]. Genera defined solely on regressive characters have been rejected in the past as shared common ancestry of regressive characters cannot be sufficiently scrutinized [41]. The genus *Anoptichthys* Hubbs & Innes, 1936, for example, has been described for the Mesoamerican (Mexico) cave fish *A. jordani* Hubbs & Innes, 1936. The genus is now considered for the aforementioned reasons a junior synonym of *Astyanax* Baird & Girard, 1854 [42,43]. Interestingly, even within the Mexican tetra *Astyanax mexicanus* (De Filippi 1853), an important cave fish model organism, eye regression and albinism were found to be convergent characters [44,45].

An additional level of complexity is represented by the lack of monophyly in *Enteromius* [20,39,46]. Yang et al. [20] suggested the presence of two poorly supported *Enteromius* clades within the African diploid small barbs, (*Enteromius* I, including *E. jae, E. hulstaerti, E. guirali*, *E. callipterus, E. camptacanthus* and *E. eburneensis*, and *Enteromius* II, including *E. trimaculatus* and *E. fasciolatus,* for which mitogenomic data are included in our phylogenetic analysis). Other studies hypothesize also the presence of two [39], but with different species, or even three lineages [40]. The taxon coverage of our analysis is too limited to draw conclusions about the different *Enteromius* groups, and further studies based on a more comprehensive taxon sampling are urgently needed to clarify the unstable taxonomy of this genus.

Furthermore, as observed in previous studies, the phylogenetic relationships within the

subfamily Smiliogastrinae remain largely unresolved [3,20] (Fig. 3). Accordingly, in our study, *Barbodes semifasciolatus* and *Chagunius chagunio* were recovered as part of the same clade in a basal position. This implies a lack of monophyly also for the genus *Barbodes* and uncertainty about the molecular classification of *B. semifasciolatus* and *C. chagunio*. This latter was recovered by Yang et al. [20] either within Smiliogastrinae (when using mitochondrial markers for phylogenetic tree reconstructions) or within Poropuntiinae (when using nuclear markers) [20,47] and it was placed as Cyprinidae *incertae sedis* by Tan and Armbruster [3]. Similarly, Yang et al. [20] recovered *Barbodes semifasciolatus* within Poropuntiinae.

As *Caecobarbus* is the oldest, next available, generic name [48] that revealed to be part of the genus *Enteromius* as presently identified, its position is of key importance for the future naming of possible subclades that would be elevated to genus rank. While waiting for additional evidence that would be needed for the taxonomic revision of these genera, the available data suggests that *Caecobarbus* might be a junior synonym of *Enteromius*. An integrative approach should now be used to properly resolve the phylogenetic position of these two genera: morphological and molecular techniques (using both nuclear and mitochondrial markers) [49–52] should be combined with a larger taxon sampling from the multiple *Enteromius* clades. As such, a multidisciplinary approach will not only allow a better understanding of the relative role of evolutionary processes such as incomplete lineage sorting or secondary introgression [53,54] but, most importantly, help to further elucidate the possible delimitation and naming of the currently identified lineages of *Enteromius*.

Acknowledgments

The authors acknowledge the financial support of the Belgian Science Policy (BELSPO) for part of this work through the Joint Experimental Molecular Unit (JEMU) and the Barcoding Facility for Organisms and Tissues of Policy Concern (BopCo). The specimen used in this research was collected in context of the Mbisa-Congo I project (2013-2018), a framework agreement project of the RMCA financed through the Belgian Development Cooperation (BDC). We want to thank the local partners Soleil Wamuini Lukayilakio (ISP-Mbanza Ngungu, DRC), Kosi Zola S. (ISP-Mbanza Ngungu, DRC) and Alexis Kimbembi (ISP-Mbanza Ngungu, DRC) for assistance in the fieldwork and the administrative support. Furthermore, we also like to thank Paul N'Lemvo (Institut Congolais pour la Conservation de la Nature, DRC), Frans Arijs (CITES Belgium) and Lore Nackaerts (Royal Museum for Central Africa) for helping with the CITES paperwork. Finally, we are grateful to Carl Vangestel (Royal Belgian Institute of Natural Sciences, JEMU) for stimulating discussions and feedback and to the two anonymous reviewers who provided feedback on an earlier version of this paper.

References:

[1] G.A. Boulenger, Description d'un poisson aveugle decouvert par M.G. Geerts dans la grotte de Thysville (Bas-Congo), Rev. Zool. Africaine. 9 (1921) 252–253.

[2] E. Vreven, A. Kimbembi ma Ibaka, S. Wamuini Lunkayilakio, The Congo blind barb: Mbanza-Ngungu's albino cave fish, in: W.R.T. Darwall, K.G. Smith, D.J. Allen, R.A. Holland, I.J. Harrison, E.G.E. Brooks (Eds.), Divers. Life African Freshwaters Underwater, under Threat. An Anal. Status Distrib. Freshw. Species throughout Mainl. Africa, 2011: pp. 74–75.

[3] M. Tan, J.W. Armbruster, Phylogenetic classification of extant genera of fishes of the order Cypriniformes (Teleostei: Ostariophysi), Zootaxa. 4476 (2018) 6–39. doi:10.11646/zootaxa.4476.1.4.

[4] R. Van der Laan, R. Fricke, Eschmeyer's Catalog of Fishes: family-group names, (2020).

[5] R. Fricke, W.N. Eschmeyer, R. Van der Laan, Eschmeyer's Catalog of Fishes: genera, species, references, (2020).

[6] J. Parzefall, Behavioural ecology of cave-dwelling fishes, in: T.J. Pitcher (Ed.), Behav. Teleost Fishes, Springer, Boston, MA, 1986: pp. 433–434.

[7] A. Ercolini, R. Berti, Morphology and response to light of *Uegitglanis zammaranoi* Gianferrari, anophthalmic phreatic fish from somalia, Monit. Zool. Ital. Suppl. 9 (1977) 183–199. doi:10.1080/03749444.1977.10736848.

[8] H. Wilkens, Genes, modules and the evolution of cave fish, Heredity (Edinb). 105 (2010) 413–422. doi:10.1038/hdy.2009.184.

[9] Convention on International Trade in Endangered Species of Wild Fauna and Flora - *Caecobarbus geertsii*, (n.d.). https://cites.org/eng/node/20215 (accessed 8 November 2019).

[10] A. Brown, R. Abell, Ecoregion 80. Thysville Caves, in: M.L. Thieme, R.A. Abell, M.L.J. Stiassny, P. Skelton, B. Lehner, G.G. Teugels, E. Dinerstein, A.K. Toham, N. Burgess, D. Olson (Eds.), Freshw. Ecoregions Africa Madagascar A Conserv. Assess., Island Press, Washington, D. C., 2005: pp. 97–98.

[11] G.S. Myers, Preface to any future classification of the cyprinid fishes of the genus *Barbus*, Stanford Ichthyol. Bull. 7 (1960) 212–215.

[12] L.K. Oellermann, P.H. Skelton, Hexaploidy in yellowfish species (*Barbus*, Pisces, Cyprinidae) from southern Africa, J. Fish Biol. 37 (1990) 105–115.

[13] A.S. Golubtsov, E.Y. Krysanov, Karyological study of some cyprinid species from Ethiopia. The ploidy differences between large and small *Barbus* of Africa, J. Fish Biol. 42 (1993) 445–455.

[14] J.F. Guégan, P. Rab, A. Machordom, I. Doadrio, New evidence of hexaploidy in 'large' African *Barbus* with some consideration on the origin of hexaploidy, J. Fish Biol. 47 (1995) 192–198.

[15] C.S. Tsigenopoulos, P. Kasapidis, P. Berrebi, Phylogenetic relationships of hexaplid large-sized barbs (genus *Labeobarbus*, Cyprinidae) based on mt DNA data, Mol. Phylogenet. Evol. 56 (2010) 851–856.

[16] P. Berrebi, A. Chenuil, P. Kotlík, A. Machordom, C.S. Tsigenopoulos, Disentangling the evolutionary history of the genus *Barbus sensu lato*, a twenty years adventure, Profr. Carlos Almaça - Estado Da Arte Em Áreas Cient. Que Desenvolveu. (2014) 29–55. doi:10.13140/2.1.1307.8080.

[17] E. Vreven, T. Musschoot, J. Snoeks, U.K. Schliewen, The African hexaploid Torini (Cypriniformes: Cyprinidae): review of a tumultuous history, Zool. J. Linn. Soc. 177 (2016) 231–305.

[18] A. Vervoort, Karyotype of *Caecobarbus geertsi* Boulenger, (Teleostei: Cyprinidae), Nucleus. 23 (1980) 76–77.

[19] E.D. Cope, Supplement on some new species of American and African fishes, Trans. Am. Philos. Soc. (New Ser. 13 (1867) 400–407.

[20] L. Yang, T. Sado, M. Vincent Hirt, E. Pasco-Viel, M. Arunachalam, J. Li, X. Wang, J. Freyhof, K. Saitoh, A.M. Simons, M. Miya, S. He, R.L. Mayden, Phylogeny and polyploidy: Resolving the classification of cyprinine fishes (Teleostei: Cypriniformes), Mol. Phylogenet. Evol. 85 (2015) 97–116. doi:10.1016/j.ympev.2015.01.014.

[21] M.J. Heuts, N. Leleup, La géographie et l'ecologie grottes du Bas-Congo. Les habitats de *Caecobarbus* Blgr., Sci. Zool. 35 (1954).

[22] B. Bushnell, BBMap: A fast, Accurate, Splice-Aware Aligner, (2014) 1–2.

[23] M. Bernt, A. Donath, F. Jühling, F. Externbrink, C. Florentz, G. Fritzsch, J. Pütz, M. Middendorf, P.F. Stadler, MITOS: Improved de novo metazoan mitochondrial genome annotation, Mol. Phylogenet. Evol. 69 (2013) 313–319. doi:10.1016/j.ympev.2012.08.023.

[24] S. Ratnasingham, P.D.N. Hebert, A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System, PLoS One. 8 (2013). doi:10.1371/journal.pone.0066213.

[25] M. Kottelat, The fishes of the inland waters of Southeast Asia: A catalogue and core bibliography of the fishes known to occur in freshwaters, mangroves and estuaries, 2013.

[26] W.N. Eschmeyer, J.D. Fong, Species by Family/Subfamily in the Catalog of Fishes, Cat. Fishes. (2018). http://researcharchive.calacademy.org/research/ichthyology/catalog/SpeciesByFamily.asp (accessed 8 April 2020).

[27] R. Froese, D. Pauly, FishBase, World Wide Web Electron. Publ. (2019). www.fishbase.org (accessed 16 December 2019).

[28] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice., Nucleic Acids Res. 22 (1994) 4673–80.

[29] R. Lanfear, B. Calcott, S.Y.W. Ho, S. Guindon, PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses, Mol. Biol. Evol. 29 (2012) 1695–1701. doi:10.1093/molbev/mss020.

[30] F. Ronquist, J.P. Huelsenbeck, MrBayes 3: Bayesian phylogenetic inference under mixed models, Bioinformatics. 19 (2003) 1572–1574. doi:DOI: 10.1093/bioinformatics/btg180.

[31] S. Tavaré, Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences, Lect. Math. Life Sci. 17 (1986) 57–86.

[32] M. Hasegawa, H. Kishinu, T. Yano, Dating of the Human-Ape Splitting by a Molecular Clock of Mitochondrial DNA, J. Mol. Evol. 22 (1985) 160–174.

[33] A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, Bioinformatics. 30 (2014) 1312–1313. doi:10.1093/bioinformatics/btu033.

[34] A. Rambaut, FigTree v1.4.3, (2016). http://tree.bio.ed.ac.uk/software/figtree/.

[35] M.A. Miller, W. Pfeiffer, T. Schwartz, Creating the CIPRES Science Gateway for inference of large phylogenetic trees, 2010 Gatew. Comput. Environ. Work. GCE 2010. (2010). doi:10.1109/GCE.2010.5676129.

[36] R. Arai, Fish Karyotypes. A check list, Springer, Tokyo, 2011.

[37] M.J. Heuts, Ecology, variation and adaptation of the blind African cave fish *Caecobarbus*, Ann. La Société R. Zool. Belgique. 82 (1951) 155–230.

[38] C.M.J. Breder, Cave fish evolution, Evolution (N. Y). 7 (1953) 179–181.

[39] M.M. Hayes, J.W. Armbruster, The Taxonomy and Relationships of the African Small

Barbs (Cypriniformes: Cyprinidae), Copeia. 105 (2017) 348–362. doi:10.1643/ci-15-348.

[40] M. Protas, W.R. Jeffery, Evolution and development in cave animals: From fish to crustaceans, Wiley Interdiscip. Rev. Dev. Biol. 1 (2012) 823–845. doi:10.1002/wdev.61.

[41] H. Wilkens, U. Strecker, Evolution in the dark. Darwin's loss without selection, Springer, 2017.

[42] R.E. Reis, S.O. Kullander, C.J. Ferraris, Pontifícia Universidade Católica do Rio Grande do Sul. Museu de Ciências e Tecnologia., Check list of the freshwater fishes of South and Central America, EDIPUCRS, 2003.

[43] A. Romero, Scientists prefer them blind: The history of hypogean fish research, Environ. Biol. Fishes. 62 (2001) 43–71. doi:10.1023/A:1011830329016.

[44] T.E. Dowling, D.P. Martasian, W.R. Jeffery, Evidence for multiple genetic forms with similar eyeless phenotypes in the blind cavefish, *Astyanax mexicanus*, Mol. Biol. Evol. 19 (2002) 446–455. doi:10.1093/oxfordjournals.molbev.a004100.

[45] H. Wilkens, Regressive evolution: ontogeny and genetics of cavefish eye rudimentation, Biol. J. Linn. Soc. 92 (2007) 287–296. doi:10.1111/j.1095-8312.2007.00840.x.

[46] Q. Ren, R.L. Myeden, Molecular phylogeny and biogeography of African diploid barbs, '*Barbus*', and allies in Africa and Asia (Teleostei: Cypriniformes), Zool. Scr. 45 (2016) 642–649.

[47] L. Yang, M. Vincent Hirt, T. Sado, M. Arunachalam, R. Manickam, K.L. Tang, A.M. Simons, H.H. Wu, R.L. Mayden, M. Miya, Phylogenetic placements of the barbin genera *Discherodontus*, *Chagunius*, and *Hypselobarbus* in the subfamily Cyprininae (Teleostei: Cypriniformes) and their relationships with other barbins, Zootaxa. 40 (2012) 26–40. doi:10.11646/zootaxa.3586.1.5.

[48] ICZN, International Code of Zoological Nomenclature, Fourth Edi, The International Trust for Zoological Nomenclature 1999, London, UK, 1999.

[49] M.K. Fujita, A.D. Leaché, F.T. Burbrink, J.A. McGuire, C. Moritz, Coalescent-based species delimitation in an integrative taxonomy, Trends Ecol. Evol. 27 (2012) 480–488. doi:10.1016/j.tree.2012.04.012.

[50] Benoit Dayrat, Towards integrative taxonomy, Biol. J. Linn. Soc. 85 (2005) 407–415.

[51] J.M. Padial, A. Miralles, I. De la Riva, M. Vences, The integrative future of taxonomy, Front. Zool. 7 (2010) 1–14. doi:10.1186/1742-9994-7-16.

[52] L.C. Gomes, T.C. Pessali, N.G. Sales, P.S. Pompeu, D.C. Carvalho, Integrative taxonomy detects cryptic and overlooked fish species in a neotropical river basin, Genetica. 143 (2015) 581–588.

[53] R.C. Schmidt, H.L. Bart, W.D. Nyingi, Multi-locus phylogeny reveals instances of mitochondrial introgression and unrecognized diversity in Kenyan barbs (Cyprininae: Smiliogastrini), Mol. Phylogenet. Evol. 111 (2017) 35–43. doi:10.1016/j.ympev.2017.03.015.

[54] B.S. Meyer, M. Matschiner, W. Salzburger, Disentangling Incomplete Lineage Sorting and Introgression to Refine Species-Tree Estimates for Lake Tanganyika Cichlid Fishes., Syst. Biol. 66 (2017) 531–550. doi:10.1093/sysbio/syw069.

Supplementary Material

Supplementary Material SM 1.

Mitochondrial genome of *Caecobarbus geertsii*. Start, stop and size of each gene (bp), inter-genic region (number of bp between two different genes, negative numbers indicate genes overlaps in bp).

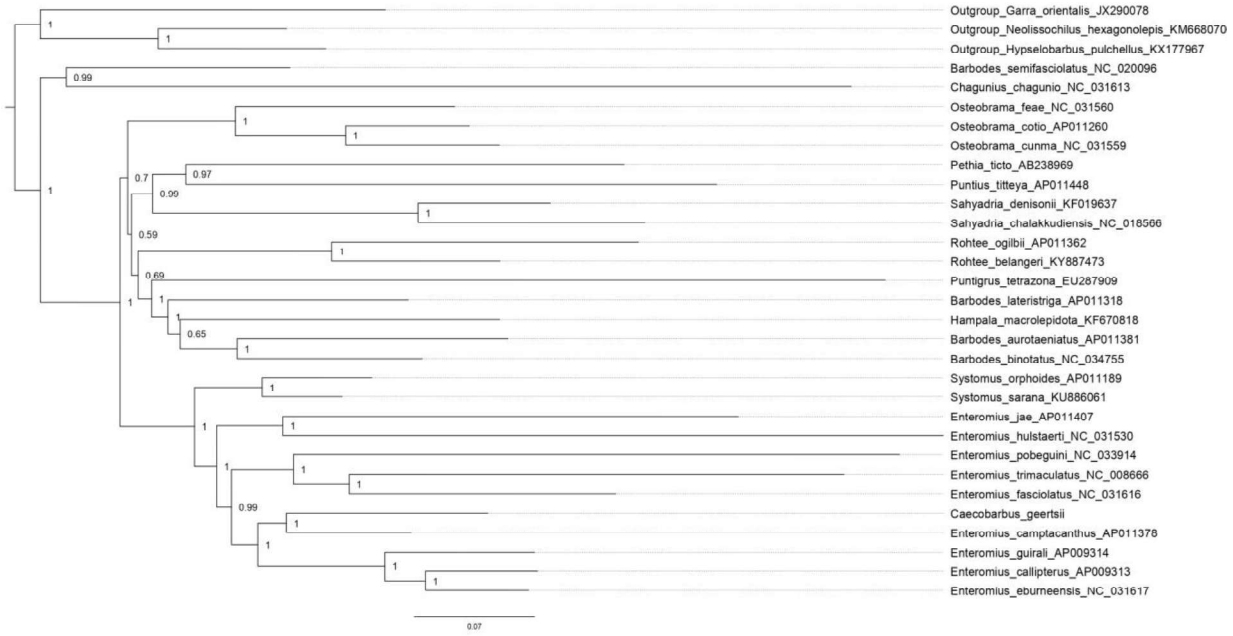
Caecobarbus geertsii

Gene	Start	End	Size	Direction	Inter-genic	Start/Stop
tRNA-Phe (trnF)	1	69	69	forward	/	
12S rRNA	70	1024	955	forward	/	
tRNA-Val (trnV)	1027	1098	72	forward	2	
16S rRNA	1099	2784	1686	forward	/	
tRNA-Leu (trnL)	2785	2862	78	forward	/	
NAD1	2863	3837	975	forward	/	ATG/TAG
tRNA-Ile (trnI)	3842	3915	74	forward	4	
tRNA-Gln (trnQ)	3913	3984	72	reverse	-3	
tRNA-Met (trnM)	3985	4053	69	forward	/	
NAD2	4054	5098	1045	forward	/	ATG/T(AA)
tRNA-Trp (trnW)	5099	5168	70	forward	/	
tRNA-Ala (trnA)	5170	5240	71	reverse	1	
tRNA-Asn (trnN)	5241	5313	73	reverse	/	
rep_region	5314	5346	33		/	
tRNA-Cys (trnC)	5347	5415	69	reverse	/	
tRNA-Tyr (trnY)	5415	5484	70	reverse	-1	
COX1	5486	7036	1551	forward	1	GTG/TAA
tRNA-Ser (trnS)	7037	7107	71	reverse	/	
tRNA-Asp (trnD)	7110	7181	72	forward	2	
COX2	7192	7882	691	forward	10	ATG/T(AA)
tRNA-Lys (trnK)	7883	7959	77	forward	/	
atp8	7961	8125	165	forward	1	ATG/TAA
atp6	8119	8802	684	forward	-7	ATG/TAA
COX3	8802	9585	784	forward	-1	ATG/T(AA)
tRNA-Gly (trnG)	9586	9658	73	forward	/	
NAD3	9659	10007	349	forward	/	ATG/T(AA)
tRNA-Arg (trnR)	10008	10077	70	forward	/	
NAD4L	10078	10374	297	forward	/	ATG/TAA
NAD4	10368	11748	1381	forward	-7	ATG/T(AA)
tRNA-His (trnH)	11749	11817	69	forward	/	
tRNA-Ser (trnS)	11818	11886	69	forward	/	
tRNA-Leu (trnL)	11888	11960	73	forward	1	
NAD5	11964	13787	1824	forward	3	ATG/TAA
NAD6	13784	14305	522	reverse	-4	ATG/TAG
tRNA-Glu (trnE)	14306	14374	69	reverse	/	
CytB	14380	15520	1141	forward	5	ATG/T(AA)
tRNA-Thr (trnT)	15521	15592	72	forward	/	
tRN-Pro (trnP)	15590	15662	73	reverse	-3	
Control region	15663	16565	903	forward	/	

Supplementary Material SM 2

Bayesian tree obtained from the analysis of 13 concatenated mitochondrial PCGs and 2 rRNA of 28 representatives of the subfamily Smiliogastrinae (including *C. geertsi*), and *Garra orientalis* (subfamily Labeoninae) *Neolissochilus hexagonolepis* and *Hypselobarbus pulchellus* (subfamily Torinae) as outgroup for tree reconstruction. Bayesian posterior probability (PP) is indicated at each node.

.



Supplementary Material SM 3

Maximum Likelihood (ML) tree obtained from the analysis of 13 concatenated mitochondrial PCGs and 2 rRNA of 28 representatives of the subfamily Smiliogastrinae (including *C. geertsi*), and *Garra orientalis* (subfamily Labeoninae) *Neolissochilus hexagonolepis* and *Hypselobarbus pulchellus* (subfamily Torinae) as outgroup for tree reconstruction. Bootstrap support is indicated at each node.

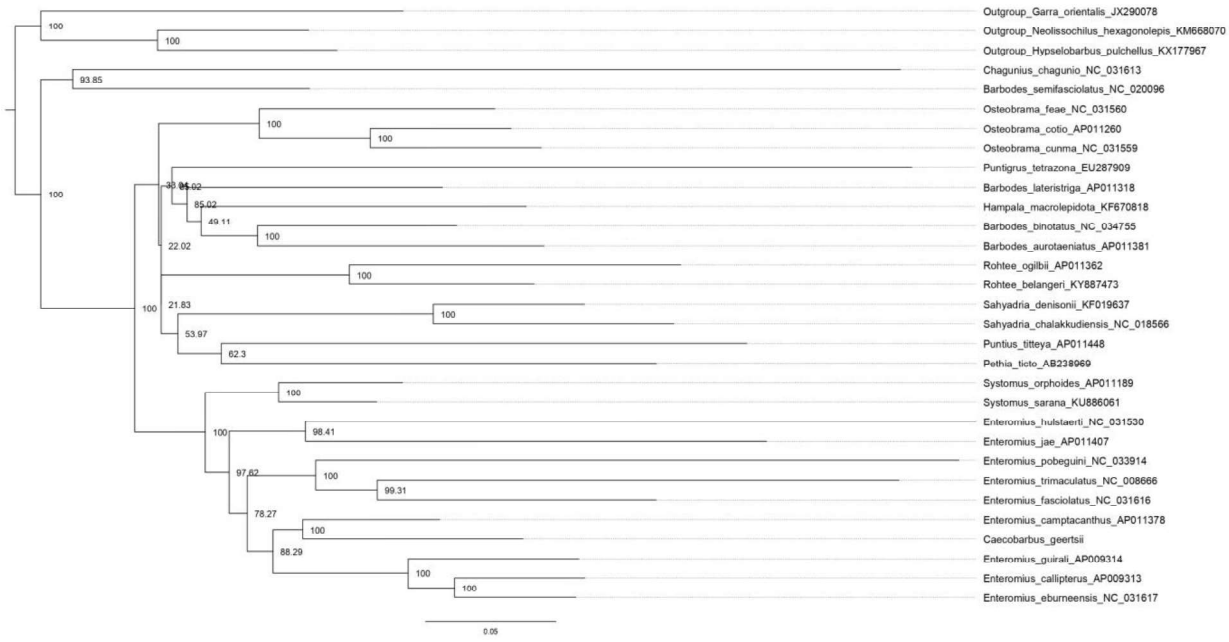


Table 1: Genome size (bp), base composition (%) and GC content (%) for the mitochondrial genome *of Caecobarbus geertsii* and of 15 publicly available mitochondrial genomes used during reference alignment. Family and subfamily are indicated following Tan and Armbruster [3].

Species	Family - Subfamily	Genome size (bp)	%A	%T	%C	%G	%CG	Reference
Caecobarbus geertsii	Cyprinidae - Smiliogastrinae	16,565	33.7	26.7	25.1	14.6	39.6	This study
Schizothorax taliensis	Cyprinidae - Schizothoracinae	16,578	29.6	25.3	27.1	17.9	45.1	Li J.T. and Rui X. (2018)
Semilabeo obscurus	Cyprinidae – Labeoninae	16,598	31.9	25.4	27.0	15.7	42.7	Liu T. and Li S. (2018)
Garra Salweenica	Cyprinidae - Labeoninae	16,960	32.5	25.5	26.7	15.2	41.9	Xiong H. and Wang X. (2017)
Gnathopogon nicholsi	Gobionidae - Sarcocheilichthyinae	16,606	29.7	27.2	25.7	17.4	43.1	He H. et al. (2017)
Sinocyclocheilus multipunctatus	Cyprinidae - Cyprininae	16,586	31.3	25.9	26.5	16.4	42.8	Zhang R. and Wang X. (2017)
Rectoris posehensis	Cyrpinidae- Labeoninae	16,594	32.1	25.9	26.5	15.5	42.0	Xiong H. and Wang X. (2013)
Herzensteinia microcephalus	Cyrpinidae - Labeoninae	16,726	28.4	27.2	26.1	18.4	44.4	He D. et al (2017)
Epalzeorhynchos frenatus	Cyprinidae - Labeoninae	16,457	32.5	26.8	25.6	15.0	40.6	Miya M. (2017)
Discogobio longibarbatus	Cyprinidae - Labeoninae	16,594	31.8	26.1	26.4	15.7	42.1	Zheng L. and Yang J. (2017)
Rohtee belangeri	Cyprinidae - Smiliogastrinae	16,602	33.1	27.6	24.3	14.9	39.3	Barman A.S. et al. (2017)
Rasbora argyrotaenia	Danioninae, Rasborinae	16,740	33.4	26.1	25.4	15.0	40.4	Kusuma W.E. et al., (2017)
Xenocypris yunnanensis	Xenocyprididae - Xenocypridinae	16,630	31.3	25.4	27.2	16.1	43.3	Xue S. (2017)
Candidia barbatus	Xenocyprididae - Opsariichthyinae	16,608	30.3	26.9	26.1	16.7	42.9	Huang,S. et al. (2017)
Microphysogobio kiatingensis	Gobionidae - Gobioninae	16,603	30.8	26.1	26.6	16.5	43.1	Zou,Y. et al. (2018)
Acheilognathus omeiensis	Acheilognathidae - /	16,774	29.0	27.7	26.0	17.3	43.3	Zou,Y. et al. (2018)

Table 2: Genome size (bp), GC content (%), genbank accession and reference for the mitochondrial genomes of *Caecobarbus geertsii*, 27 species from the subfamily Smiliogastrinae and *Neolissochillus hexagonolepis*, *Hypselobarbus pulchellus* (subfamily Torinae) and *Garra orientalis* (subfamily Labeoninae) used as outgroup for tree reconstruction. Genome lengths indicated with * lack the D-loop.

Species	species author	length	GC%	Genbank accession	Reference
Caecobarbus geertsii	(Boulenger, 1921)	16,565	39.6	MK457704	this study
Barbodes aurotaeniatus	(Tirant, 1885)	16,562	42.2	AP011381	Miya M. (2009)
Barbodes binotatus	(Valenciennes, 1842)	16,573	43	KY305681	Wu C.L. et al. (2017)
Barbodes lateristriga	(Valenciennes, 1842)	16,586	41.6	AP011318	Miya M. (2009)
Barbodes semifasciolatus	(Günther, 1868)	16,594	41.7	KC113209	Jang-Liaw N.H. et al. (2013)
Chagunius chagunio	(Hamilton, 1822)	17,017	42.2	AP011373	Miya M. (2009)
Enteromius callipterus	(Boulenger, 1907)	16,859	38.4	AP009313	Saitoh K. et al. (2006)
Enteromius camptacanthus	(Bleeker, 1863)	15,665*	40.0	AP011378	Miya M. (2009)
Enteromius eburneensis	(Poll, 1941)	16,678	38.3	AP011379	Miya M. (2009)
Enteromius fasciolatus	(Günter, 1868)	16,566	29.4	AP011377	Miya M. (2009)
Enteromius guirali	(Thominot, 1886)	16,793	39.1	AP009314	Saitoh K. et al. (2006)
Enteromius hulsaerti	(Poll, 1945)	16,775	37.5	AP011194	Miya M. (2009)
Enteromius jae	(Boulenger, 1903)	15,660*	38.9	AP011407	Miya M. (2009)
Enteromius pobeguini	(Pellegrin, 1911)	16,993	42.0	AP012061	Miya M. (2009)
Enteromius trimaculatus	(Peters, 1852)	16,417	39.3	NC_008666	Saitoh K. et al. (2006)
Garra orientalis	(Nichols, 1925)	17,288	41.7	JX290078	Su et al. (2013)
Hampala macrolepidota	(Kuhl & van Hasselt, 1823)	16,766	41.8	KF670818	Liu M. and Liu S. (2013)
Hypselobarbus pulchellus	(Day, 1870)	16,588	43.7	KX177967	Sahoo et al. (2016)
Neolissochilus hexagonolepis	(McClelland, 1839)	16,563	43.1	KM668070	Goel et al. (2014)
Osteobrama cotio	(Hamilton 1822)	16,584	41.6	AP011260	Miya M. (2009)
Osteobrama cunma	(Day, 1888)	16,650	41.1	AP011261	Miya M. (2009)
Osteobrama feae	(Vinciguerra, 1890)	16,578	44.4	AP011262	Miya M. (2009)
Pethia ticto	(Hamilton, 1822)	17,302	40	AB238969	Saitoh K. et al. (2006)
Puntigrus tetrazona	(Bleeker, 1855)	16,550	40.2	EU287909	Nguyen J.V. et al. (2007)
Puntius titteya	(Deraniyagala 1929)	15,776*	40.1	AP011448	Miya M. (2009)
Rohtee belangeri	(Valenciennes, 1844)	16,602	39.3	KY887473	Barman A.S. et al. (2017)
Rohtee ogilbii	(Sykes, 1839)	15,682*	41.7	AP011362	Miya M. (2009)
Sahyadria chalakkudiensis	(Menon, Rema Devi & Thobias, 1999)	16,989	40.1	JX311437	Joseph T.C. et al. (2012)
Sahyadria denisonii	(Day, 1865)	16,899	41.4	KF019637	Gopalakrishnan A. et al. (2013)
Systomus orphoides	(Valenciennes, 1842)	16,593	41.1	AP011189	Miya M. (2009)
Systomus sarana	(Hamilton, 1822)	16,590	41.4	KU886061	Biswal J.R. et al. (2016)

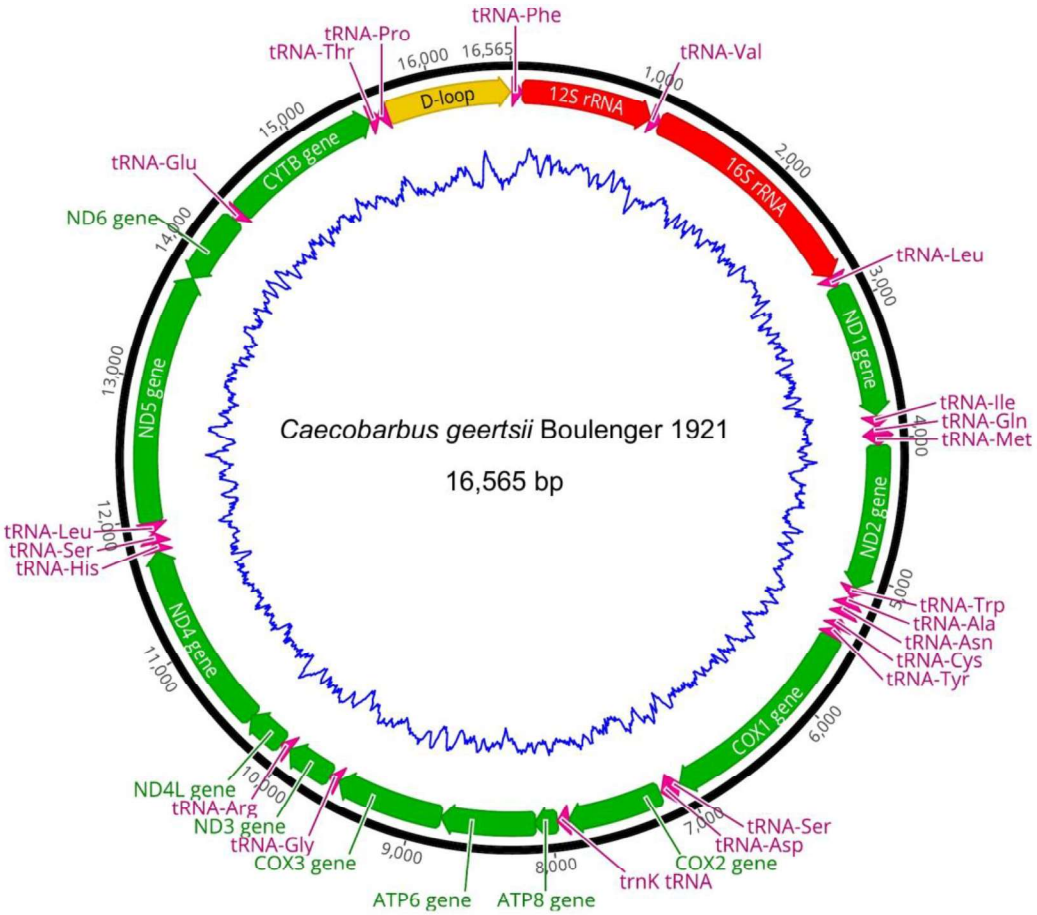
Figure Captions

Figure 1. Schematic representation of the complete mitogenome of the *C. geertsii*. Base pair numbering (black circle), protein coding genes (in green), rRNAs (in red), and tRNAs (in fuchsia) and the GC skew (blue line) are indicated.

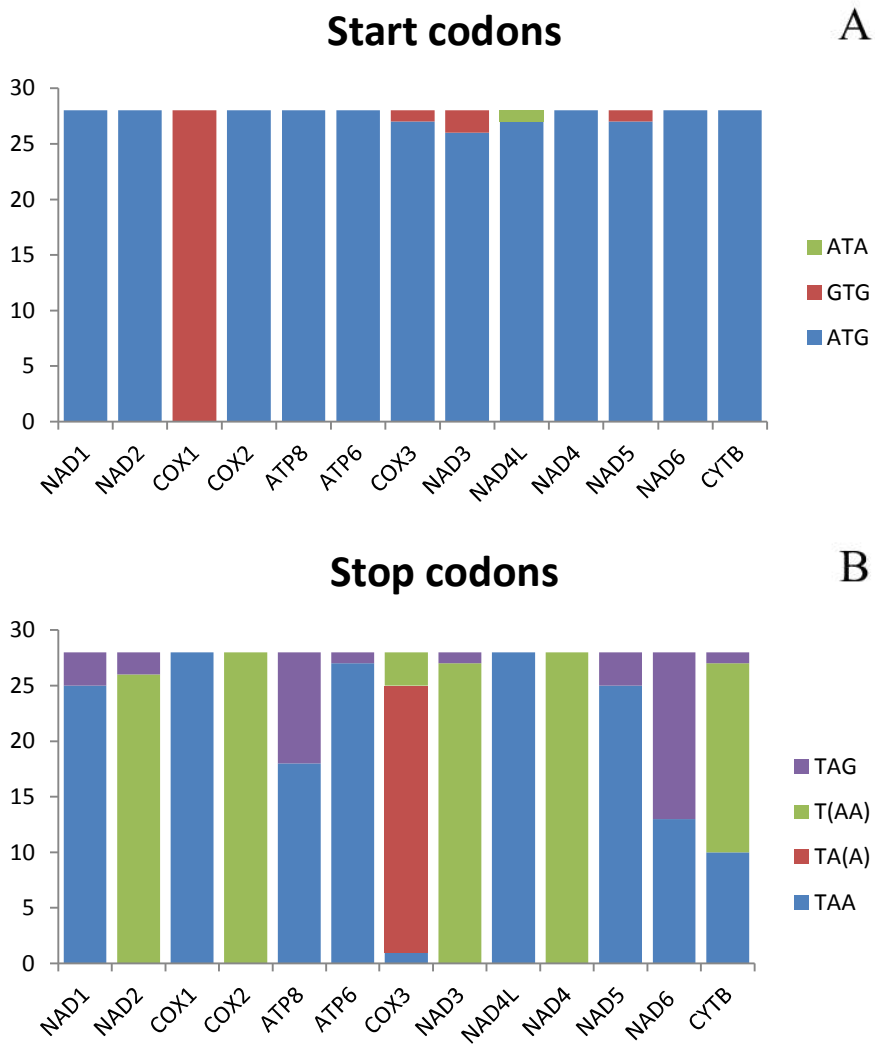
Figure 2: Start and stop codons of 28 members of Smiliogastrinae (including *C. geertsii*). See Table 2 for complete list) .

Figure 3: Maximum Likelihood (ML) and Bayesian trees obtained from the analysis of 13 concatenated mitochondrial PCGs and 2 rRNA of 28 representatives of the subfamily Smiliogastrinae (including *C. geertsi*), and *Garra orientalis* (subfamily Labeoninae) *Neolissochilus hexagonolepis* and *Hypselobarbus pulchellus* (subfamily Torinae) as outgroup for tree reconstruction. For each node (separated by slashes): Bayesian PP and ML bootstrap are indicated (-: node not recovered). Not supported nodes (PP < 0.95 / bootstrap < 75%) are indicated in light-grey. In green *Caecobarbus geertsii*, in blue two *Enteromius* species assigned to *Enteromius* Lineage II by Yang et al. [20].

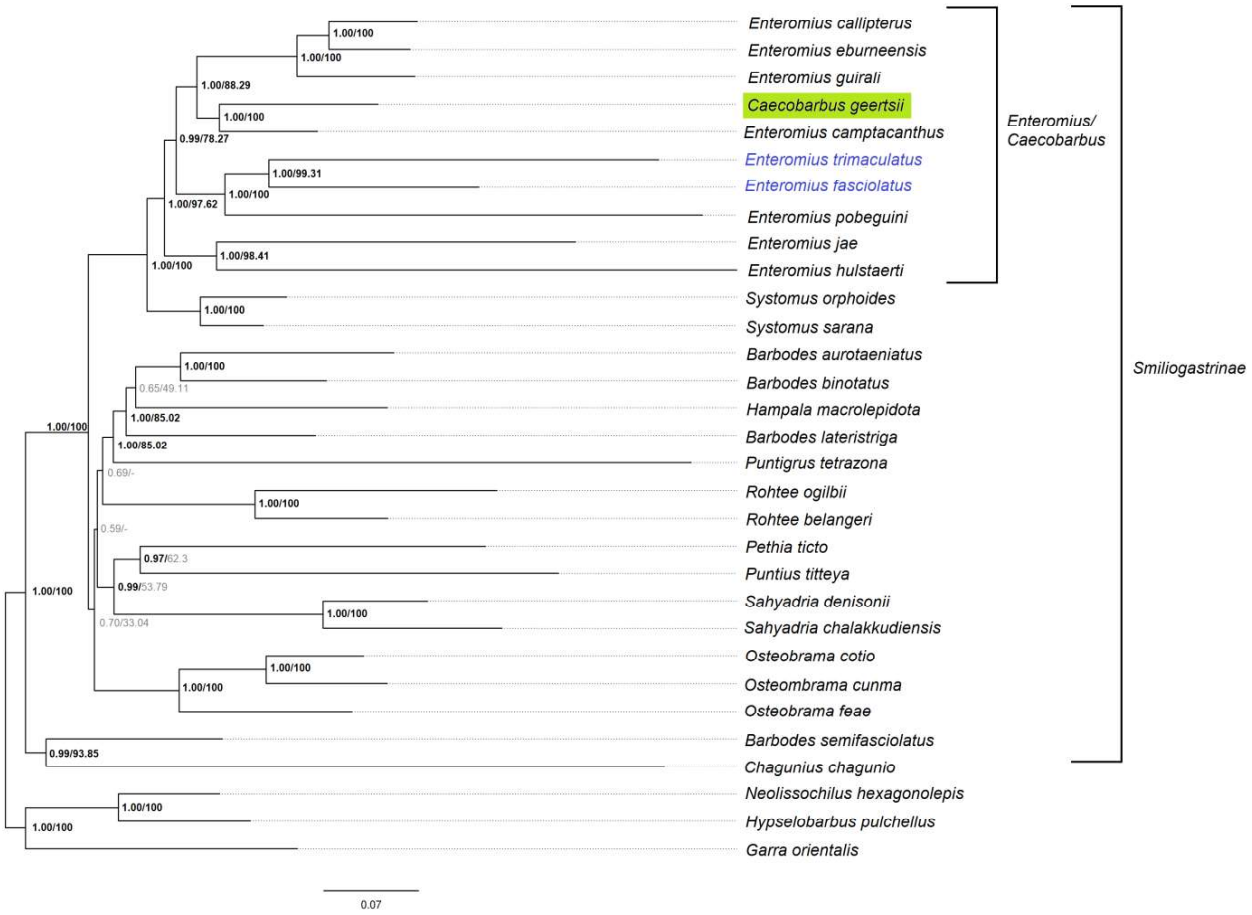
International Journal of Biological Macromolecules, N. Mullens, G. Sonet, E. Decru, M. Virgilio, J. Snoeks, E. Vreven. Fig. 1



International Journal of Biological Macromolecules, N. Mullens, G. Sonet, E. Decru, M. Virgilio, J. Snoeks, E. Vreven. Fig 2



International Journal of Biological Macromolecules, N. Mullens, G. Sonet, E. Decru, M. Virgilio, J. Snoeks, E. Vreven. Fig 3



Author statement

Nele Mullens: Conceptualization, Methodology, Investigation, Formal analysis, Writing- Original draft preparation **Gontran Sonet:** Methodology, Investigation, Writing-Reviewing and Editing **Eva Decru:** Writing-Reviewing and Editing **Massimiliano Virgilio:** Conceptualization, Funding acquisition, Supervision, Writing-Reviewing and Editing **Snoeks Jos:** Writing-Reviewing and Editing **Emmanuel Vreven:** Resources, Writing-Reviewing and Editing