Submitted: 24 Oct. 2019
Accepted: 6 Jan. 2020
Editor: A. Dettaï

The complete mitochondrial genome of *Thymallus thymallus* (Linnaeus, 1758) (Actinopterygii, Salmonidae) obtained by long range PCRs and double multiplexing

by

Gaël P.J. DENYS* (1, 2), Giulia SECCI-PETRETTO (3, 4)
& André GOMES-DOS-SANTOS (3, 4)

Résumé. – Le génome mitochondrial complet du *Thymallus thymallus* (Linnaeus, 1758) (Actinopterygii, Salmonidae) obtenu par PCRs longues et double multiplexage.

Le génome mitochondrial complet de *Thymallus thymallus* (Actinopterygii, Salmonidae) provenant d'un voucher enregistré en collection et capturé dans l'Ain (bassin du Rhône) a été séquencé. La séquence d'une longueur de 16 660 pb, contient 13 gènes codants, 22 ARN de transfert dont 2 ARNt-Leu et 2 ARNt-Ser, 2 gènes ARN ribosomiques et la région de contrôle. Un espace intergénique et des séquences se chevauchant ont été trouvés. La composition des bases du mitogénome est de 27,5% A, 27,9% T, 27,1% C et 17,5% G. Ce mitogénome a été séquencé pour moins de 20 euros.

Key words. – Mitogenome – *Thymallus thymallus* – Long PCR – Sample multiplexing.

Graylings (*Thymallus* genus) belong to the monogeneric Thymallinae subfamily of Salmonidae and display an almost circumpolar distribution with at least 19 species listed (Dyldin *et al*., 2017; Persat *et al*., 2019). *Thymallus thymallus* (Linnaeus, 1758) is a European species with a widespread distribution, ranging from England in the west to the Ural Mountains in the east and Montenegro in the south to northern Finland and Barent Sea tributaries in Russia in the north (Persat, 1996). In France, this species is native in the Meuse, Rhine and Rhône river drainages (Persat *et al*., 2019).

Although several mitogenomes of grayling species have become available in the last years (e.g. Yasuike *et al*., 2010; Ma *et al*., 2016), all were obtained through PCR amplification (17 reactions) followed by Sanger sequencing. This approach has revealed itself to be time and resource consuming with an estimated cost of around 90 euros per sample. Moreover, no voucher material is linked to the sequences, which is problematic for taxonomic work, especially when new species are being described (e.g. Persat *et al*., 2019). Without vouchers, proper assignment of taxa to mitogenome sequences is difficult.

In this paper, we describe the mitogenome of *Thymallus thymallus* from a specimen conserved in a Museum collection, amplified and sequenced according to the multiplexing protocol of Hinsinger *et al*. (2015).

MATERIAL AND METHODS

Voucher

The vouchered specimen is from the Zoological Collections of the Claude Bernard Lyon 1 University (UCBLZ) and recorded in the catalogue UCBLZ 2012.9.971. The specimen (OMB11) was caught in the Ain River (Rhône drainage) at Marigny (Jura Dept.) on August 31, 2015, by Henri Persat. It was morphologically identified by Henri Persat according to Persat et al. (2019).

Brief material and method

The mitogenome was obtained from a fin-clip stored in 95% ethanol. DNA extraction was carried on an EpMotion Robot using MN Biomedical extraction kits, according to the manufacturer protocols. Amplification, PCR amplifications and sequencing were done according to Hinsinger *et al*. (2015) with three long 6 kbp overlapping PCRs (primers in Tab. I). The time of initial denaturation and denaturation of each cycle was reduced to 20 s and 10 s,

Table I. – Primers for the long PCRs used to amplify the mitogenome of *Thymallus thymallus*.

Long PCR	Primer name	5'-3' sequence	Source
Mt1	12S-L1091R	AAACTGGGATTAGATACCCCACTAT	Kocher *et al*. (1989)
	MtH7061 Salmo	GTGTTATGCGGTTGGCTTGAAAC	This study
Mt2	MtL5231 Salmo	TAGGTGGGAAGGCCTCGATCCTACA	
	MtH11944 Salmo	CATAGCTGCTACTTGGATTTGCACCA	
Mt3	MtL11910 Salmo	CAGCTCATCCGTTGGTCTTAGGAAC	
	12S-H1478 Salmo	AACTTGGGGAGAGTGACGGGCGGTGTGT	

(1) Unité Mixte de Service Patrimoine Naturel – Centre d'expertise et de données (UMS 2006 AFB – CNRS – MNHN), Muséum national d'Histoire naturelle, 36 rue Geoffroy Saint-Hilaire CP 41, 75005 Paris, France. [gael.denys@mnhn.fr]
(2) Laboratoire de Biologie des organismes et écosystèmes aquatiques (BOREA), MNHN, CNRS, IRD, SU, UCN, UA, 57 rue Cuvier CP26, 75005 Paris, France.
(3) CIIMAR, Interdisciplinary Centre of Marine and Environmental Research of the University of Porto, Terminal de Cruzeiros do Porto de Leixões, Avenida General Norton de Matos, S/N, 4450-208 Matosinhos, Portugal. [lasecci@hotmail.it] [andrepousa64@gmail.com]
(4) Department of Biology, Faculty of Sciences, University of Porto, Rua do Campo Alegre 1021/1055, Porto, Portugal.
* Corresponding author

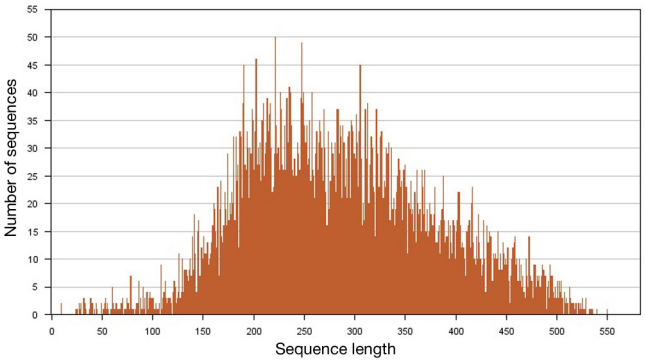

Figure 1. – Sequence lengths for 7760 nucleotide sequences allowing the assembling of the *Thymallus thymallus* mitogenome .

respectively, in the PCR cycles. The reads were assembled using a reference sequence (GenBank accession number FJ853655) as bait according to Hahn *et al*. (2013) with the Geneious 11.2.2 software (Kearse *et al*., 2012). The consensus sequence was controlled and annotated using MitoAnnotator (Iwasaki *et al*., 2013).

Sequence quality

7760 reads (average: 285.6, SD: 94.4, range: 10-550; Fig. 1), obtained from the 3 long range PCRs, were assembled.

The mitogenome read coverage ranged between 26 and 379, with a mean read coverage of 134.

Phylogenetic reconstruction

For phylogenetic analysis, additional mitogenomes belonging to eight distinct *Thymallus* species, plus a sequence of *Coregonus lavaretus* (Linnaeus, 1758), *Hucho taimen* (Pallas, 1773) and *Brachymystax lenok* (Pallas, 1773) were retrieved from GenBank.

The 13 protein-coding genes were isolated, individually aligned in Bioedit (Hall, 1999) and subsequently concatenated using a Fasta Alignment Joiner tool in FaBox 1.5 (Villesen, 2007) online platform. The concatenated dataset included 12 sequences, reaching a total length of 11433 bp.

The best evolutionary model was inferred to be GTR+G+I in jModelTest (Darriba *et al*., 2012) for both Akaike and Bayesian information criterion. Phylogenetic analysis was inferred by Maximum Likelihood (ML) using RAxML-HPC2 Workflow on XSEDE (version 8.2.10) (Stamatakis, 2014) with the inferred substitution model and 1000 bootstrap iterations on the CIPRES Science Gateway (Miller *et al*., 2010) online platform.

Nucleotide diversity and sequence divergence between the newly sequenced mitogenome and the *T. thymallus* sample retrieved from GenBank were calculated in Geneious 11.2.2.

RESULTS

Sequence description and phylogenetic analysis

The newly obtained mitogenome has a total length of 16660 bp and includes 13 protein-coding genes, 22 transfer RNA genes including 2 tRNA-Leu and 2 tRNA-Ser, two ribosomal RNA genes and a control region following the standard vertebrate order (Tab. II). Intergenic spaces and overlapping sequences were found. Six coding genes have an incomplete codon stop: NADH2, COII, ATP6, NADH3, NADH4 and Cyt b. The base composition of the entire genome was 27.5% for A, 27.9% for T, 17.5% for G and 27.1% for C.

The ML phylogeny obtained shows high support values for all nodes (Fig. 2). The *Thymallus* sequences are separated in two monophyletic groups. The first one (including the newly sequenced mitogenome) comprises the four nominal species: *T. thymallus*, *T. brevirostris*, *T. arcticus*, *T. baicalolenensis*. The second includes the three nominal species *T. tugarinae*, *T. grubii*, *T. yaluensis*. Furthermore, the new sequence clusters with the *T. thymallus* sequence from Finland (GenBank Accession number FJ853655) with a divergence of 1% (pairwise p-distance) and 130 differences.

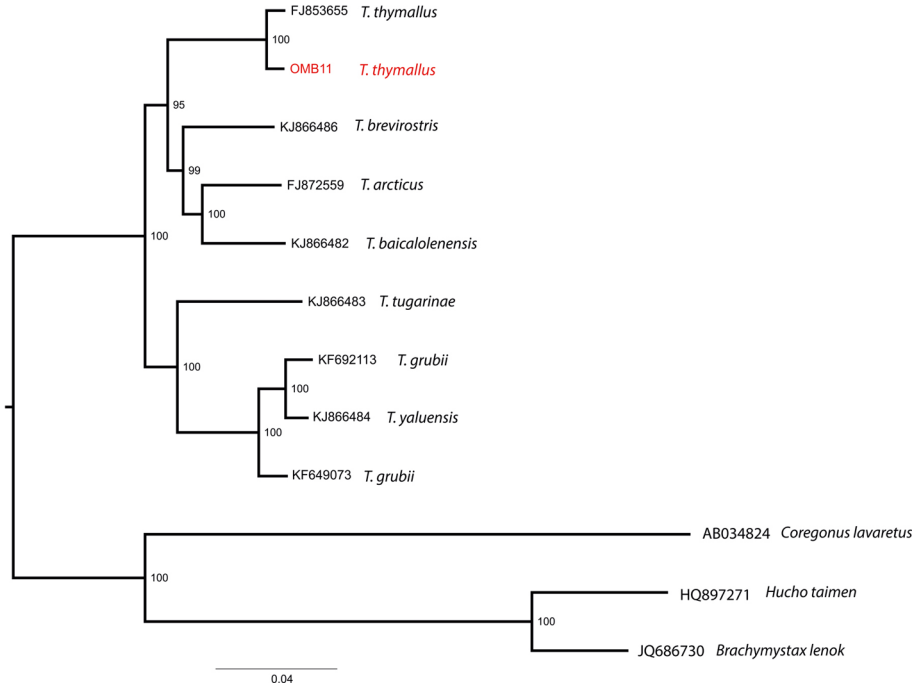

Figure 2. – Maximum Likelihood phylogenetic tree inferred with the 13 protein coding genes. The values of bootstrap are represented beside the nodes.

DISCUSSION

Comparison to already available genomes

The gene order was conserved in all 9 *Thymallus* mitogenomes present in this analysis. The ML phylogeny is consistent with a recent study, which also relied on mitogenomics (Ma *et al*., 2016). Furthermore, the phylogenetic reconstruction (Fig. 2), by clustering the new mitogenome OMB 11 with the other *T. thymallus* sequence, as well as the low pairwise p-distance (1%) observed between the two sequences, supports identification of the voucher specimen.

Comparison to other sequencing methods

The method described by Hinsinger *et al*. (2015) for amplifying mitogenomes using 3 long PCRs and double multiplexing for sequencing yields a whole mitochondrial genome for less than 20 euros of sequencing costs because the PCR products for this sequence represent only 0.4% of the whole DNA inserted in the sequencing chip. This strategy is substantially cheaper than shotgun sequencing methods (280 euros according to Murienne *et al*. (2016)). Furthermore, it is less demanding in terms of laboratory procedures and minimizes the potential amplicon mix-up that can occur with Sanger sequencing of shorter amplicons (Salas *et al*., 2005). However, this method is only applicable on fresh and non-degraded material.

Acknowledgements. – This work was supported by the Muséum national d'Histoire naturelle (MNHN), the UMS PatriNat 2006, the laboratory BOREA 7208 and the Agence Française pour la Biodiversité (AFB). This work was funded by the partnership MNHN-AFB 2016-2018 and partially supported by the Portuguese Foundation for Science and Technology (FCT) under grant SFRH/BD/137935/2018 (AGS) and SFRH/BD/139069/2018 (GSP). We are grateful to Nicolas Poulet for supporting the MNHN-AFB partnership, Henri Persat who sent us the fin clip and Hanae Essalmi for her help for the mitogenome assembling. Laboratory access and assistance was provided by the "Service de Systématique Moléculaire" of the Muséum national d'Histoire naturelle (CNRS UMS 2700). Finally, we thank warmly Steven Weiss for checking English and improving the manuscript.

Table II. – Genes composition of the complete mitochondrial genome of *Thymallus thymallus* (GenBank Accession Number MN852233) including their position, length and direction.

Name	Type	First nucleotide	Last nucleotide	Length (bp)	Direction
tRNA-Phe	tRNA	1	68	68	Forward
12S rRNA	rRNA	69	1014	946	Forward
tRNA-Val	tRNA	1015	1086	72	Forward
16S rRNA	rRNA	1087	2765	1679	Forward
tRNA-Leu	tRNA	2766	2840	75	Forward
NADH1 gene	gene	2841	3815	975	Forward
tRNA-Ile	tRNA	3824	3895	72	Forward
tRNA-Gln	tRNA	3893	3963	71	Reverse
tRNA-Met	tRNA	3963	4031	69	Forward
NADH2 gene	gene	4032	5080	1049	Forward
tRNA-Trp	tRNA	5082	5151	70	Forward
tRNA-Ala	tRNA	5154	5222	69	Reverse
tRNA-Asn	tRNA	5224	5296	73	Reverse
tRNA-Cys	tRNA	5330	5396	67	Reverse
tRNA-Tyr	tRNA	5397	5467	71	Reverse
COI gene	gene	5469	7019	1551	Forward
tRNA-Ser	tRNA	7020	7091	72	Reverse
tRNA-Asp	tRNA	7095	7167	73	Forward
COII gene	gene	7182	7872	691	Forward
tRNA-Lys	tRNA	7873	7946	74	Forward
ATP8 gene	gene	7948	8115	168	Forward
ATP6 gene	gene	8106	8788	683	Forward
COIII gene	gene	8789	9574	786	Forward
tRNA-Gly	tRNA	9574	9643	70	Forward
NADH3 gene	gene	9644	9992	349	Forward
tRNA-Arg	tRNA	9993	10062	70	Forward
NADH4L gene	gene	10063	10359	297	Forward
NADH4 gene	gene	10353	11733	1381	Forward
tRNA-His	tRNA	11734	11802	69	Forward
tRNA-Ser	tRNA	11803	11872	70	Forward
tRNA-Leu	tRNA	11874	11946	73	Forward
NADH5 gene	gene	11947	13785	1839	Forward
NADH6 gene	gene	13782	14303	522	Reverse
tRNA-Glu	tRNA	14304	14372	69	Reverse
CYTB gene	gene	14377	15517	1141	Forward
tRNA-Thr	tRNA	15518	15589	72	Forward
tRNA-Pro	tRNA	15589	15658	70	Reverse
control region D-loop	D-loop	15659	16660	1002	Forward

REFERENCES

DARRIBA D., TABOADA G. L., DOALLO R. & POSADA D., 2012. – jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods*, 9(8): 772. DOI: 10.1038/nmeth.2109

DYLDIN Y.V., HANEL L., ROMANOV V.I. & PLESNÍK J., 2017. – A review of the genus *Thymallus* (Pisces: Salmoniformes, Salmonidae, Thymallinae) with taxonomic notes. *Bull. Lampetra*, 8: 103-126.

HAHN C., BACHMANN L. & CHEVREUX B., 2013. – Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads – a baiting and iterative mapping approach. *Nucl. Acids Res.*, 41: e129. DOI: 10.1093/nar/gkt371

HALL T.A., 1999. – BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.*, 41: 95-98.

HINSINGER D.D., DEBRUYNE R., THOMAS M., DENYS G.P.J., MENNESSON M., UTGE M. & DETTAI A., 2015. – Fishing for barcodes in the Torrent: from COI to complete mitogenomes on NGS platforms. *DNA Barcodes*, 3: 170-186. DOI: 10.1515/dna-2015-0019

IWASAKI W., FUKUNAGA T., ISAGOZAWA R., YAMADA K., MAEDA Y., SATOH T.P., SADO T., MABUCHI K., TAKESHIMA H., MIYA M. & NISHIDA M., 2013. – MitoFish and MitoAnnotator: A mitochondrial genome database of fish with an accurate and automatic annotation pipeline. *Mol. Biol. Evol.*, 30: 2531-2540. DOI: 10.1093/molbev/mst141

KEARSE M., MOIR R., WILSON A., STONES-HAVAS S., CHEUNUNG M., STURROCK S., BUXTON S., COOPER A., MARKOWITZ S., DURAN C., THIERER T., ASHTON B., MEINTJES P. & DRUMMOND A., 2012. – Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28: 1647-1649. DOI: 10.1093/bioinformatics/bts199

KOCHER T.D., THOMAS W.K., MEYER A., EDWARDS S.V., PÄÄBO S., VILLABLANCA F.X. & WILSON A.C., 1989. - Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci. USA*, 86: 6196-6200. DOI: 10.1073/pnas.86.16.6196

MA B., JIANG H., SUN P., CHEN J., LI L., ZHANG X. & YUAN L., 2015. – Phylogeny and dating of divergences within the genus *Thymallus* (Salmonidae: Thymallinae) using complete mitochondrial genomes. *Mitochondrial DNA A*, 27(5): 3602-3611. DOI: 10.3109/19401736.2015.1079824

MILLER M. A., PFEIFFER W. & SCHWARTZ T., 2010. – Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *In*: 2010 Gateway Computing Environments Workshop (GCE). Ieee, p. 1-8. DOI: 10.1109/GCE.2010.5676129

MURIENNE J., JEZIOSKI C., HOLOTA H., COISSAC E., BLANCHET S. & GRENOUILLET G., 2016. – PCR-free shotgun sequencing of the stone loach mitochondrial genome (*Barbatula barbatula*). *Mitochondrial DNA A*, 27(6): 4211-4212. DOI: 10.3109/19401736.2015.1022744

PERSAT H., 1996. – Threatened populations and conservation of the European grayling, *Thymallus thymallus* (L. 1758). *In*: Conservation of Endangered Freshwater Fish in Europe (Kirchhofer A. & Hefti D., Eds). ALS Advances in Life Sciences, Birkhäuser Basel: 233-247. DOI: 10.1007/978-3-0348-9014-4_23

PERSAT H., WEISS S., FROUFE E., SECCI-PETRETTO G. & DENYS G.P.J., 2019. – A third European species of grayling (Actinopterygii, Salmonidae), endemic to the Loire River basin (France), *Thymallus ligericus* n. sp. *Cybium*, 43(3): 233-238. DOI: 10.26028/cybium/2019-433-004

ROZAS J., FERRER-MATA A., SÁNCHEZ-DELBARRIO J.C., GUIRAO-RICO S., LIBRADO P., RAMOS-ONSINS S.E. & SÁNCHEZ-GRACIA A., 2017. – DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.*, 34(12): 3299-3302. DOI: 10.1093/molbev/msx248

SALAS A., YAO Y.G., MACAULAY V., VEGA A., CARRACEDO A. & BANDELT H.J., 2005. – A critical reassessment of the role of mitochondria in tumorigenesis. *PLoS Med.*, 2(11): e296. DOI: 10.1371/journal.pmed.0020296

STAMATAKIS A., 2014. – RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9): 1312-1313. DOI: 10.1093/bioinformatics/btu033

VILLESEN P., 2007. – FaBox: an online toolbox for FASTA sequences. *Mol. Ecol. Notes*, 7(6): 965-968. DOI: 10.1111/j.1471-8286.2007.01821.x

YASUIKE M., JANTZEN S., COOPER G.A., LEDER E., DAVIDSON W.S. & KOOP B.F., 2010. – Grayling (Thymallinae) phylogeny within salmonids: complete mitochondrial DNA sequences of *Thymallus arcticus* and *Thymallus thymallus*. *J. Fish Biol.*, 76: 395-400. DOI: 10.1111/j.1095-8649.2009.02494.x