The first Margaritiferidae male (M-type) mitogenome: Gene order as a potential character for determining higher-order phylogeny within Unionida, Bivalvia.

This submission is intended as a Research Note

MANUEL LOPES-LIMA[1,2], MIGUEL FONSECA[1,2,3], DAVID C. ALDRIDGE[4], ARTHUR BOGAN[5], HANG MING GAN[6,7], MOHAMED GHAMIZI[8], RONALDO SOUSA[9], AMÍLCAR TEIXEIRA[10], SIMONE VARANDAS[11], DAVID ZANATTA[12], ALEXANDRA ZIERITZ[13], AND ELSA FROUFE[1]

[1]*CIIMAR/CIMAR - Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Terminal de Cruzeiros do Porto de Leixões, Av. General Norton de Matos s/n, 4450-208 Matosinhos, Portugal.*

[2]*CIBIO/InBIO - Research Center in Biodiversity and Genetic Resources, Universidade do Porto, Campus Agrário de Vairão, Rua Padre Armando Quintas, 4485-661 Vairão, Portugal.*

[3]Department of Biochemistry, Genetics and Immunology, University of Vigo, Vigo 36310, Spain.

[4]*Aquatic Ecology Group, The David Attenborough Building, Department of Zoology, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, United Kingdom*

[5]*Research Laboratory, North Carolina State Museum of Natural Sciences, MSC 1626, Raleigh, NC 27699-1626, USA.*

[6]*Monash University Malaysia Genomics Facility, Monash University Malaysia, Jalan Lagoon Selatan, Bandar Sunway, 47500 Petaling Jaya, Selangor Malaysia.*

[7]*School of Science, Monash University Malaysia, Jalan Lagoon Selatan, Bandar Sunway, 47500 Petaling Jaya, Selangor, Malaysia.*

[8]*Muséum d'Histoire Naturelle de Marrakech, Université Cadi Ayyad, Faculté des Sciences, Semlalia, B.P. 2390 Marrakech, Morocco.*

[9]*CBMA - Centre of Molecular and Environmental Biology, Department of Biology, University of Minho, Campos de Gualtar, 4710-057 Braga, Portugal.*

[10]*CIMO-ESA-IPB – Mountain Research Centre, School of Agriculture, Polytechnic Institute of Bragança, Campus de Santa Apolónia, Apartado 1172, 5301-854 Bragança, Portugal.*

[11]*CITAB-UTAD – Centre for Research and Technology of Agro-Environment and Biological Sciences, University of Trás-os-Montes and Alto Douro, Apartado 1013, 5001-811 Vila Real, Portugal.*

[12]*Biology Department, Institute for Great Lakes Research, Central Michigan University, Biosciences 2408, Mount Pleasant, MI 48859, USA.*

[13]*School of Environmental and Geographical Sciences, Faculty of Science, University of Nottingham Malaysia Campus, Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia.*

Short Running Head: *MARGARITIFERA MAROCANA* MITOGENOMES AND UNIONIDA PHYLOGENY

Correspondence: M. Lopes-Lima; manuelpmlopeslima@gmail.com

The family Margaritiferidae (Bivalvia: Unionida), comprising 12 extant species, is widely distributed across the northern hemisphere of North America, Europe and Asia (Bolotov *et al*., 2016). Most species in this family have dramatically declined over the last century, with nine out of the 12 Margaritiferidae species assessed as threatened in the most recent IUCN Red List (IUCN, 2016). Among these is the Moroccan pearl mussel *Margaritifera marocana* (Pallary, 1918), considered one of the 100 most threatened species on the planet (Baillie & Butcher, 2012). The species is now restricted to two small streams in the Oum Er Rbia and Sebou basins, and conservation measures are urgently needed (Lopes-Lima, M. pers. obs.).

Beyond the conservation concern, Unionida are also biologically interesting. They present an unusual mechanism of mitochondrial inheritance called Doubly Uniparental Inheritance (DUI), in which all individuals have the typical maternally transmitted mtDNA (F-type), but the males possess in their germ cells a paternally inherited mtDNA instead (M-type) (Zouros *et al*., 1994; Plazzi *et al*., 2015). So far, DUI has been observed in species from three families of Unionida, i.e., Unionidae, Hyriidae and Margaritiferidae. However, to date, no whole M-type mitogenome has been published for either species belonging to the latter two families.

The gene arrangement within mitogenomes is conserved in many taxonomic groups. For example, most vertebrates share the same gene order (Pereira *et al*., 2000). In other groups, like Bivalvia, the mitochondrial genome arrangement is more variable, although not many distinct gene orders have been described so far (Plazzi *et al*., 2013). Due to this fact, mitogenome rearrangements seem to be rare events that are difficult to replicate. In this context, mitogenome gene order might be used as an additional character for phylogenetic inference. However, its utility for Bivalvia and particularly in the Unionida phylogeny has never been tested.

The order Unionida has six recognized families with around 800 species (Lopes-Lima *et al*., 2014). However, the phylogenetic relationships among these families have been highly contentious with the available studies presenting conflicting results (*e.g.*, Bogan & Hoeh, 2000; Graf & Cummings, 2006). This lack of coherence among studies has been consistently attributed to the low number of molecular markers and insufficient taxon sampling (Bogan & Roe, 2008; Fonseca *et al*., 2016).

Under the above mentioned assumptions the aims of the present study are to: i) sequence and analyse the whole M- and F-type mitogenomes of *Margaritifera marocana*, ii) infer the phylogenetic relationships among Unionida species using all both the F- and M-type mtDNA sequences publicly available, and iii) determine the gene order of all analysed mitogenomes and evaluate its phylogenetic utility.

One male specimen deposited in the Muséum d'Histoire Naturelle de Marrakech (Voucher MHNM16ZMB23) from the Laabid River (GPS WGS84: 32.142334, -7.027595) was dissected for gonadal and mantle tissue collections. DNA extractions followed Froufe *et al.* (2016). The complete M- and F-type mitogenomes of *M. marocana* were then sequenced, assembled and annotated using an established pipeline (Gan *et al*., 2014). The F and M mitogenomes have been deposited in GenBank database under the accession numbers KY131953 and KY131954, respectively.

The two newly obtained *M. marocana* mitogenome sequences were aligned with all (43) M- and F-type Unionida mitogenome sequences available on GenBank as of March 2016, as well as with the F- and M-type mitogenomes of *Mytilus galloprovincialis* as outgroups (list of genomes and respective accession numbers used supplied on request). DNA (NUC) and amino acid (AA) sequences of all mtDNA protein-coding genes (PCG) except ATP8 and the gender-specific open reading frames (M-ORF, H-ORF and F-ORF) were used in the phylogenetic analyses. The sequences of each gene were aligned using MAFFT software (version 7.304, Katoh & Standley, 2013) and trimmed with GUIDANCE (version 1.5, Penn *et al*., 2010; see Froufe *et al*., 2016 for the parameters used). The gene alignments were then concatenated, resulting in two alignments with the following length: 14,350 aligned nucleotide positions or 6,246 aligned amino acid + nucleotide positions (4,085 aligned amino acids positions and 2,161 aligned nucleotide positions from the rRNAs genes). The optimal partitioning scheme (i.e. the best set of non-overlapping partitions that cover the whole alignment) for each alignment was selected using PartitionFinder v. 1.1.1 software (Lanfear *et al*. 2012) under the greedy algorithm with proportional branch lengths across partitions. The best substitution models of DNA and protein evolution for each partition were selected under BIC ranking method (Schwarz, 1978). The codon positions of the protein-coding genes and each rRNA were defined as the initial data blocks for the partitioning schemes search. The Maximum Likelihood (ML) phylogenetic inference was performed using RAxML (v. 8.0.0, Stamatakis, 2014) with 100 rapid bootstrap

replicates and 20 ML searches. The Bayesian methodology (BI) was applied using MrBayes v. 3.2.1 (Ronquist *et al*., 2012) with two independent runs ($1x10^7$ generations with a sampling frequency of 1 tree for every 100 generations), each with four chains (3 hot and 1 cold). All runs reached convergence (average standard deviation of split frequencies below 0.01). The posterior distribution of trees was summarized in a 50% majority rule consensus tree (burn-in of 25%).

The length of the two newly sequenced mitogenomes of *M. marocana*, 16,001 bp for the female haplotype and 17,562 bp for the male haplotype, is within the typical range for each gender specific haplotype within Unionida. The sequenced haplotypes include the 13 protein-coding genes typically found in metazoan mitochondrial genomes, the gender-specific ORF described for all Unionida mitogenomes with DUI system (Breton *et al*., 2009, 2011a), and 22 transfer RNA (tRNA) and 2 ribosomal RNA (rRNA) genes. The M-type genome is the largest sequenced to date within the Unionida. M-type genomes are generally larger than F-type genomes due to the larger size of the protein-coding genes COX2 and M-ORF in M-type genomes compared to the orthologous genes in the F-type genomes (Breton *et al*., 2009). The four intergenic regions identified in the *M. marocana* M-type genome were analysed in the search for the M-ORF. The results of the blast search (Altschul *et al*., 1997) retrieved a significant hit with another Margaritiferidae M-ORF (*Margaritifera monodonta*, evalue= 4e-34), and the Fickett test score of 1.201 (Fickett, 1982; a score > 0.95 means the sequence is probably coding) suggesting that the M-ORF is located in the region between the genes tRNA-D and ND4L.

The M-type mitogenome of *M. marocana* presents a novel gene order within the Unionida (Fig. 1). The F-type mitogenome gene order is the same as already observed for the two previously available Margaritiferidae F-type mitogenomes (Breton *et al*., 2011b; Yang *et al*., 2015).

Figure 1 to be presented here

All the phylogenies inferred in this study support the reciprocal monophyly of both (Unionidae + Margaritiferidae) F- and M-type lineages (Fig. 2 represents the topology of the BI-NUC tree; all other phylogenetic trees figures supplied on request). Additionally, the monophyly of Unionidae (both F- and M-

type) and Margaritiferidae (F-type) and all Unionidae subfamilies are well supported in all inferred mtDNA trees, with the exception of the Unioninae, for which monophyly was only well supported in the BI-NUC tree. The remaining phylogenetic trees (BI-AA, ML-NUC and ML-AA) showed conflicting results regarding the position of the clade comprising *Arconaia lanceolata* and *Lanceolaria grayana* (Fig. 2). These conflicting results have already been found in previous studies where distinct mitogenome phylogenetic methodologies revealed distinct tree topologies (Huang *et al*., 2013; Fonseca *et al*., 2016).

Figure 2 to be presented here

Five distinct mtDNA gene orders have been detected in the present dataset, three in the F-type lineage and two in the M-type lineage. In the F-type lineage, gene order UF1 is shared by the Unionidae subfamilies Anodontinae, Ambleminae, and Unioninae, whereas gene orders UF2 and MF1 are found in the represented species of the Gonideinae subfamily and the family Margaritiferidae, respectively (Figs. 1 and 2). Mapping gene orders over the inferred mtDNA phylogeny suggests that UF1 might be ancestral within the Unionidae and UF2 derived in the ancestral lineage of the Gonideinae. However, this result has limited supporting value because no mitogenome sequences, and therefore no gene order information, are available for three out of the seven presently recognised Unionidae subfamilies. Future inclusion of mtDNA gene orders of these currently unrepresented subfamilies could change the inference of the ancestral and derived mtDNA gene orders within Unionidae evolutionary history. In the M-type, only one gene arrangement per family is obtained: UM1 for the Unionidae and MM1 for the Margaritiferidae. Due to the fact that the Unionida is a very old order (Graf & Cummings, 2007), and as a consequence of the number of distinct mitogenome gene arrangements already found, it is likely that as novel mitogenomes from additional Unionida families and subfamilies became available, more arrangements will be found. This may help to clarify the higher order phylogenetic relationships within the freshwater mussels.

ACKNOWLEDGEMENTS

This work was partially supported by the Mohamed bin Zayed Species Conservation Fund under project «Biodiversity and conservation of the critically endangered freshwater mussels in Morocco: ecogeographic, genetic and physiological information» (Reference 15256799), by the IUCN SOS Save Our Species Fund under project «Breeding the most endangered bivalve on Earth: *Margaritifera marocana*» (Reference 2015B-015) and by the Portuguese Foundation for Science and Technology (FCT) under grants SFRH/BPD/70654/2010 (MMF); SFRH/BPD/108445/2015 (EF).

REFERENCES

ALTSCHUL, S., MADDEN, T.L., SCHÄFFER, A.A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**: 3389-3402.

BAILLIE, J.E.M. & BUTCHER, E.R. 2012. *Priceless or Worthless? The World's Most Threatened Species*. Zoological Society of London, London.

BOGAN, A.E. & HOEH, W.R. 2000. On becoming cemented: evolutionary relationships among the genera in the freshwater bivalve family Etheriidae (Bivalvia: Unionoida). *Geological Society, London, Special Publications*, **177**: 159-168.

BOGAN, A.E. & ROE, K.J. 2008. Freshwater bivalve (Unioniformes) diversity, systematics, and evolution: status and future directions. *Journal of the North American Benthological Society*, **27**: 349-369.

BOLOTOV, I.N., VIKHREV, I.V., BESPALAYA, Y.V., GOFAROV, M.Y., KONDAKOV, A.V., KONOPLEVA, E.S., BOLOTOV, N.N. & LYUBAS, A.A. 2016. Multi-locus fossil-calibrated phylogeny, biogeography and a subgeneric revision of the Margaritiferidae (Mollusca: Bivalvia: Unionoida). *Molecular Phylogenetics and Evolution*, **103**: 104-121.

BRETON, S., BEAUPRE, H.D., STEWART, D.T., PIONTKIVSKA, H., KARMAKAR, M., BOGAN, A.E., BLIER, P.U. & HOEH, W.R. 2009. Comparative mitochondrial genomics of freshwater mussels (Bivalvia: Unionoida) with doubly uniparental inheritance of mtDNA: gender-specific open reading frames and putative origins of replication. *Genetics*, **183**: 1575-1589.

BRETON, S., GHISELLI, F., PASSAMONTI, M., MILANI, L., STEWART, D.T. & HOEH, W.R. 2011a. Evidence for a fourteenth mtDNA-encoded protein in the female-transmitted mtDNA of marine mussels (Bivalvia: Mytilidae). *PloS One*, **6**: e19365.

BRETON, S., STEWART, D.T., SHEPARDSON, S., TRDAN, R.J., BOGAN, A.E., CHAPMAN, E.G., RUMINAS, A.J., PIONTKIVSKA, H. & HOEH, W.R. 2011b. Novel protein genes in animal mtDNA: a new sex determination system in freshwater mussels (Bivalvia: Unionoida)? *Molecular Biology and Evolution*, **28**: 1645-1659.

FICKETT, J.W. 1982. Recognition of protein coding regions in DNA sequences. *Nucleic Acids Research*, **10**: 5303-5318.

FONSECA, M.M., LOPES-LIMA, M., EACKLES, M.S., KING, T.L. & FROUFE, E. 2016. The female and male mitochondrial genomes of *Unio delphinus* and the phylogeny of freshwater mussels (Bivalvia: Unionida). *Mitochondrial DNA B Resources*, DOI: 10.1080/23802359.2016.1241677.

FROUFE, E., GAN, H.M., LEE, Y.P., CARNEIRO, J., VARANDAS, S., TEIXEIRA, A., ZIERITZ, A., SOUSA, R. & LOPES-LIMA, M. 2016. The male and female complete mitochondrial genome sequences of the endangered freshwater mussel *Potomida littoralis* (Cuvier, 1798) (Bivalvia : Unionidae). *Mitochondrial DNA Part A*, **27**: 3571-3572.

GAN, H.M., SCHULTZ, M.B. & AUSTIN, C.M. 2014. Integrated shotgun sequencing and bioinformatics pipeline allows ultra-fast mitogenome recovery and confirms substantial gene rearrangements in Australian freshwater crayfishes. *BMC Evolutionary Biology*, **14**: 19.

GRAF, D.L. & CUMMINGS, K.S. 2006. Palaeoheterodont diversity (Mollusca: Trigonioida+ Unionoida): what we know and what we wish we knew about freshwater mussel evolution. *Zoological Journal of the Linnean Society*, **148**: 343-394.

GRAF, D.L. & CUMMINGS, K.S. 2007. Review of the systematics and global diversity of freshwater mussel species (Bivalvia: Unionoida). *Journal of Molluscan Studies*, **73**: 291-314.

HUANG, X.C., RONG, J., LIU, Y., ZHANG, M.H., WAN, Y., OUYANG, S., ZHOU, C.H. & WU, X.P. 2013. The complete maternally and paternally inherited mitochondrial genomes of the endangered

freshwater mussel *Solenaia carinatus* (Bivalvia: Unionidae) and implications for Unionidae Taxonomy. *PLoS One*, **8**: e84352.

IUCN. 2016. The IUCN Red List of Threatened Species. Version 2016-2. http://www.iucnredlist.org. Downloaded on 01 November 2016.

KATOH, K. & STANDLEY, D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, **30**: 772-780.

LANFEAR, R., CALCOTT, B., HO, S. Y. W., & GUINDON, S. 2012. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, **29**: 1695-1701.

LOPES-LIMA, M., TEIXEIRA, A., FROUFE, E., LOPES, A., VARANDAS, S. & SOUSA, R. 2014. Biology and conservation of freshwater bivalves: past, present and future perspectives. *Hydrobiologia*, **735**: 1-13.

PENN, O., PRIVMAN, E., ASHKENAZY, H., LANDAN, G., GRAUR, D. & PUPKO, T. 2010. GUIDANCE: a web server for assessing alignment confidence scores. *Nucleic Acids Research*, **38**: W23–W28.

PEREIRA, S.L. 2000. Mitochondrial genome organization and vertebrate phylogenetics. *Genetics and Molecular Biology*, **23**: 745-752.

PLAZZI, F., RIBANI, A. & PASSAMONTI, M. 2013. The complete mitochondrial genome of *Solemya velum* (Mollusca: Bivalvia) and its relationships with Conchifera. *BMC Genomics*, **14**: 409.

PLAZZI, F., CASSANO, A. & PASSAMONTI, M. 2015. The quest for Doubly Uniparental Inheritance in heterodont bivalves and its detection in *Meretrix lamarckii* (Veneridae: Meretricinae). *Journal of Zoological Systematics and Evolutionary Research*, **53**: 87-94.

RONQUIST, F., TESLENKO, M., VAN DER MARK, P., AYRES, D.L., DARLING, A., HÖHNA, S., LARGET, B., LIU, L., SUCHARD, M.A. & HUELSENBECK, J.P. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**: 539-542.

SCHWARZ, G. 1978. Estimating the dimension of a model. *The Annals of Statistics*, **6**: 461-464.

230 STAMATAKIS, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
231 phylogenies. *Bioinformatics*, **30**: 1312-1313.
232 YANG, S., MI, Z., TAO, G., LIU, X., WEI, M. & WANG, H. 2015. The complete mitochondrial genome
233 sequence of *Margaritiana dahurica* Middendorff. *Mitochondrial DNA*, **26**: 716-717.
234 ZOUROS, E., BALL A.O., SAAVEDRA, C. & FREEMAN, K.R. 1994. An unusual type of mitochondrial
235 DNA inheritance in the blue mussel *Mytilus*. *Proceedings of the National Academy of Sciences*, **91**:
236 7463–7467.

Figure 1. Diagrams of the five distinct gene orders detected on all Unionida from the present dataset. In the female F-type lineage, three gene orders are depicted: Unionidae F-type 1 (UF1), Unionidae F-type 2 (UF2), and Margaritiferidae F-type 1 (MF1). In the male M-type lineage, two gene arrangements are shown: Unionidae M-type 1 (UM1) and Margaritiferidae M-type 1 (MM1).

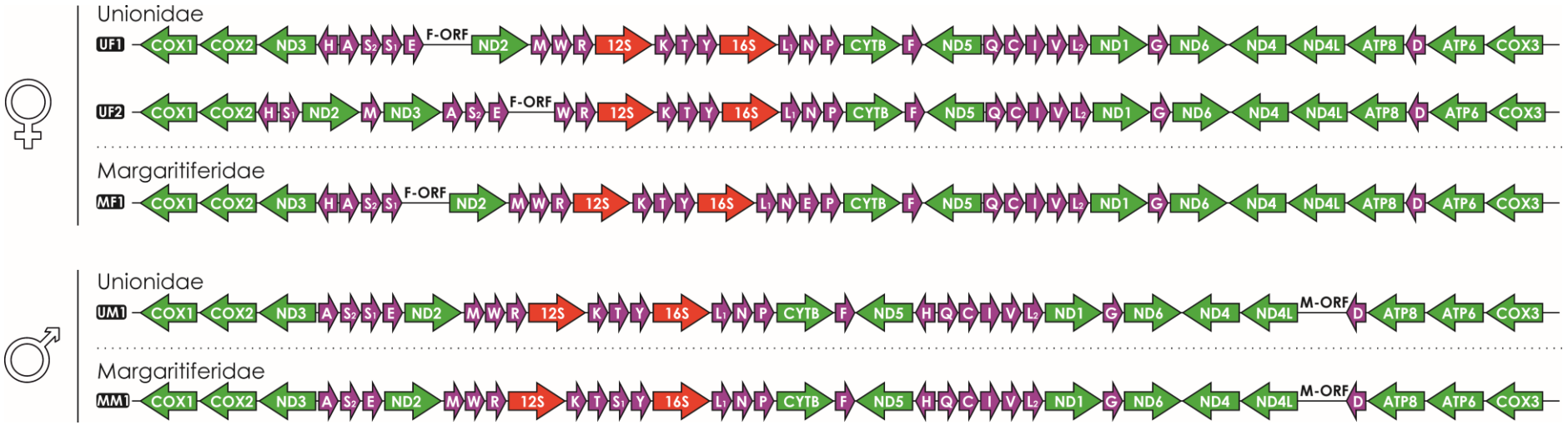
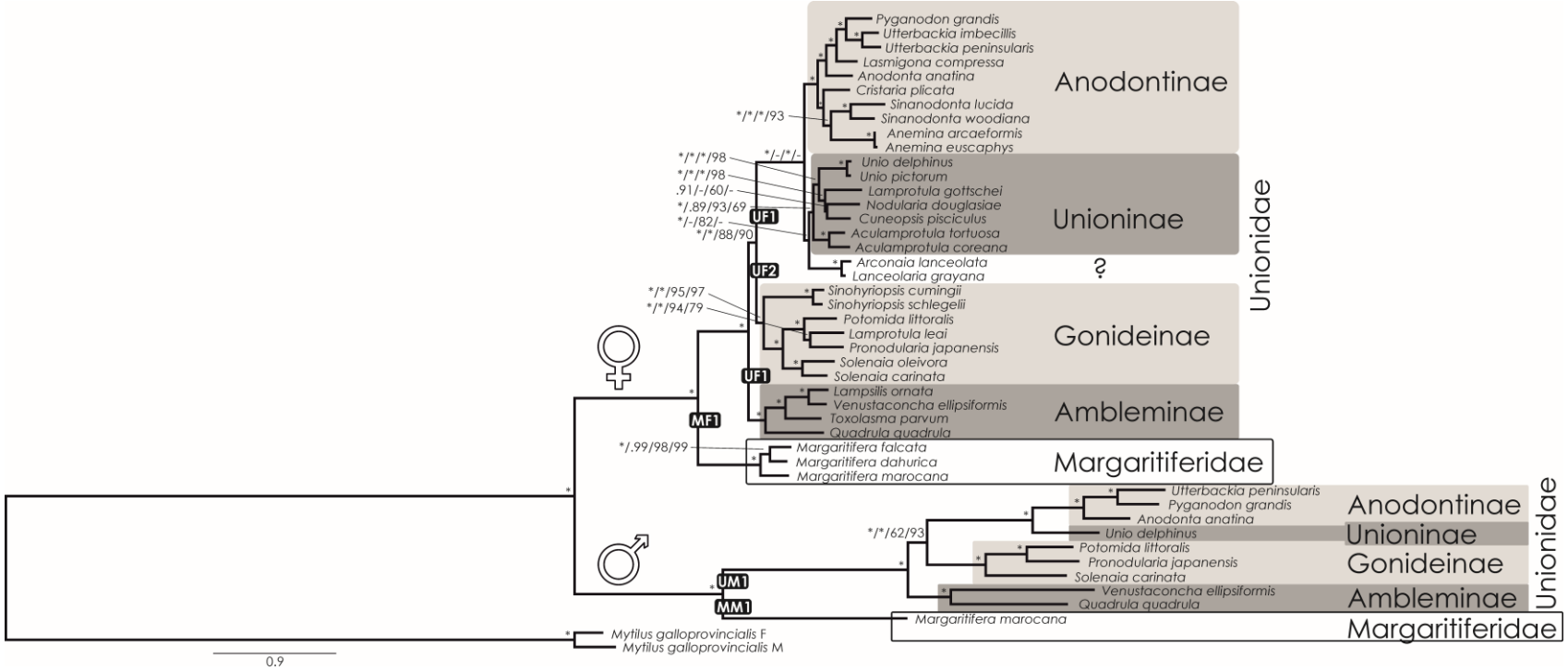


Figure 2. Phylogenetic (BI-NUC) tree of Unionida (freshwater mussels) estimated from 14 concatenated individual mtDNA gene sequences (12 protein-coding and 2 rRNA genes). The values for branch support are represented in the following order: 1) Bayesian Posterior Probabilities for BI-NUC tree, 2) Bayesian Posterior Probabilities for BI-AA tree, 3) ML bootstrap support values for ML-NUC, and 4) ML bootstrap support values for ML-AA tree. Maximum supporting values (BI = 1 and ML = 100) are represented with "*". All five distinct detected gene orders are mapped in the phylogeny branches (see Fig. 1 for gene order codes).



247

248