bioRxiv preprint doi: https://doi.org/10.1101/2021.06.19.449104; this version posted June 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Yet another mitochondrial genome of the Pacific cupped oyster: the published mitogenome of *Alectryonella plicatula* (Ostreinae) is based on a misidentified *Magallana gigas* (Crassostreinae)

Daniele Salvi[1, @], Emanuele Berrilli[1], Matteo Garzia[1], Paolo Mariottini[2]

1 Department of Health, Life and Environmental Sciences, University of L'Aquila, Via Vetoio, 67100 Coppito, L'Aquila, Italy.

2 Dipartimento di Scienze, Università Roma Tre, Viale G. Marconi 446, 00146 Rome, Italy

@: Corresponding author: Daniele Salvi, danielesalvi.bio@gmail.com

Abstract

The recently published mitochondrial genome of the fingerprint oyster *Alectryonella plicatula* (Gmelin, 1791) with GenBank accession number MW143047 was resolved in an unexpected phylogenetic position, as sister to the Pacific cupped oyster *Magallana gigas* (Thunberg, 1793) and share with this species three typical gene duplications that represent robust synapomorphies of the *Magallana* clade. In this study, we verified the identity of MW143047 using direct comparisons of single gene sequences, DNA barcoding and phylogenetic analyses. BLAST searches using each of the 12 protein coding genes and rRNA genes extracted from MW143047 as query retrieved *M. gigas* as best hit with 100% sequence identity. MW143047 is nested within the clade formed by *M. gigas* sequences, with virtually no difference between their terminal branch lengths, both in the *cox1* gene tree (based on 3639 sequences) and in the 16S gene tree (based on 1839 sequences), as well as in the Maximum Likelihood mitogenomic tree based on concatenated sequence of 12 PCGs. Our findings suggest that the original specimen used for mitogenome sequencing was misidentified and represents an individual of *M. gigas*. This study reinforces the notion that morphological shell analysis alone is not sufficient for oyster identification, not even at high taxonomic ranks such as subfamilies. While it is well established that morphological identification of oysters should be validated by molecular data, this study emphasizes that also molecular data should be taxonomically validated by means of DNA barcoding and phylogenetic analyses. The implications of the publication of taxonomically misidentified sequences and mitogenomes are discussed.

Keywords: DNA barcoding; *Magallana*; misidentification; Ostreidae; oyster; phylogeny

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.19.449104; this version posted June 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Oysters are distributed worldwide in temperate and tropical waters and several of them have a great economic importance. However, taxonomic identification of oysters based on morphological characters is challenging, even for species locally cultivated since centuries (e.g. Wang et al., 2004; Hsiao et al., 2016). Indeed, oysters' shells show a high degree of phenotypic plasticity driven by environmental factors, therefore, shell morphology is often uninformative or misleading for taxonomic identification and classification. The use of molecular data has been fruitful for species identification and has resulted in a well-established phylogeny and systematics of oysters (Salvi et al., 2014; Salvi & Mariottini, 2017). The mitochondrial genome has been the most valuable source of molecular data for oyster species identification (DNA barcoding), phylogenetic reconstruction and classification (e.g. Wang et al., 2004; Liu et al., 2011; Salvi et al., 2014; Raith et al., 2016). Moreover, mitochondrial gene rearrangements, such as transpositions and duplications, has provided additional characters for phylogenetic inference, classification and diagnosis of oysters' genera and subfamilies (Salvi & Mariottini, 2021). Molecular resources of oyster are continuously growing, and most studies currently implement these data for taxonomic identification. For this purpose, a reliable reference of taxonomically identified sequences and mitogenomes is necessary (Bortolus, 2008; Jin et al., 2020; Salvi et al., 2020).

Recently, the complete mitochondrial genome of the fingerprint oyster *Alectryonella plicatula* (Gmelin, 1791), with GenBank accession number MW143047, has been characterised (Wang et al., 2021) and resolved in an unexpected phylogenetic position, as sister to the Pacific cupped oyster *Magallana gigas* (Thunberg, 1793). Unfortunately, in this mitogenome announcement the phylogenetic position of MW143047 is described in a cladogram with arbitrary branch lengths (Wang et al., 2021), therefore masking the true evolutionary divergence between MW143047 and the mitogenome of *M. gigas* (see Botero-Castro et al., 2016). However, their sister relationship is surprising and in sharp contrast with all previous phylogenetic studies that have consistently established the placement of *A. plicatula* within the lophinae lineage, that is nested within the subfamily Ostreinae Rafinesque, 1815, whereas *M. gigas* belong to the well-defined clade of Indo-Pacific Crassostreinae Scarlato & Starobogatov, 1979 (O'Foighil & Taylor, 2000; Salvi et al., 2014; Crocetta et al., 2015; Salvi & Mariottini, 2017; Al-Kandari et al., 2021). Moreover, the newly published mitogenome MW143047 conforms to the mitochondrial gene arrangement of *M. gigas*, that is characterised by the duplication of *trnK*, *trnQ* and *rrnS* genes that are exclusive of the *Magallana* clade (Ren et al., 2010) and represent robust synapomorphies of this clade (Salvi et al., 2014; Salvi & Mariottini, 2017, 2021). These intriguing points are urgent to clarify as MW143047 might become the mitogenomic reference of *A. plicatula*. In this study, we verified the taxonomic identification of Wang et al (2021) using available quality control guidelines for taxonomic validation of new mitogenomes (Botero-Castro et al., 2016).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.19.449104; this version posted June 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Materials and Methods

We verified the identity of MW143047 using DNA barcoding and phylogenetic analyses.

We extracted from the mitogenome MW143047 the two barcoding fragments commonly used for oysters, the *cox1* and the 3' half portion of the 16S rRNA (Liu et al 2011; Crocetta et al., 2015), as well the remaining protein coding genes and rRNAs (12S and the 5' half portion of the 16S) using Geneious Prime 2021 (Biomatters Ltd., Auckland, New Zealand). Sequence of each gene were used as query in BLAST searches using default settings. Sequences of the barcoding markers *cox1* and the 16S were aligned with oysters' sequences available from public database (BOLD and NCBI) assembled, dereplicated, and aligned following the procedure by Salvi et al. (2020). A Neighbor-Joining (NJ) tree was constructed based on uncorrected *p*-distance values in MEGA v. 7 (Kumar et al., 2016) with pairwise deletion and 100 replicates of bootstrap (BS).

We inferred a Maximum Likelihood (ML) tree based on the concatenated sequences of 12 protein-coding genes (PCGs) of the same oyster taxa analysed by Wang et al. (2021) plus six additional mitogenome sequences of *M. gigas*, to further assess phylogenetic relationships and divergence between the latter and the mitogenome MW143047. ML analyses were performed in IQTREE v 1.6.12 (Nguyen et al., 2015) using for each gene partition the best substitution model determined by the ModelFinder module (Kalyaanamoorthy et al., 2017) and 1000 replicates of ultrafast bootstrapping.

Results

Results of BLAST searches using as query the *cox1* and the 16S sequences extracted from MW143047 retrieved as best hits sequences assigned to *M. gigas* with a sequence identity of 100% (sequence identity ranging from 99.85% to 100% among the best 10 hits for *cox1* and of 100% for 16S; Table 1). The same result was obtained in BLAST searches using as query the other 11 protein coding genes and rRNAs extracted from MW143047, with 100% of nucleotides identical to multiple sequences of *M. gigas*.

In the gene tree based on 3639 *cox1* sequences (Fig 1a) and in the gene tree based on 1839 16S sequences (Fig 1b) MW143047 clustered with *M. gigas* with maximum bootstrap support (BS=100%).

In the ML mitogenomic tree (Fig 2) MW143047 is nested within the clade formed by *M. gigas* sequences, with virtually no difference between their terminal branch lengths. This clade was sister to the mitogenome sequence of *M. angulata* (BS=100%) within the well supported clade formed by *Magallana* species (BS=100%).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.19.449104; this version posted June 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1. Top ten best hits of BLAST results using as query the sequences of the barcoding fragments *cox1* (above) and 16S rRNA (below) extracted from the complete mitochondrial genome MW143047.

Query sequence: *cox1* MW143047									
Accession	**Reported scientific name**	**Current scientific name**	**Isolate / Voucher**	**Max Score**	**Total Score**	**Query Cover**	**E-value**	**% Identity**	**Accession Lenght**
MN862563	*Crassostrea gigas*	*Magallana gigas*	isolate EU1	1205	1205	69%	0	100.00%	655
KJ855245	*Crassostrea gigas*	*Magallana gigas*	isolate WF34	1205	1736	100%	0	100.00%	18225
KJ855244	*Crassostrea gigas*	*Magallana gigas*	isolate YK05	1205	1736	100%	0	100.00%	18225
KJ855241	*Crassostrea gigas*	*Magallana gigas*	isolate CgJap23	1205	1736	100%	0	100.00%	18225
FJ717608	*Crassostrea gigas*	*Magallana gigas*	voucher LBDM385	1205	1205	69%	0	100.00%	692
HM626169	*Crassostrea gigas*	*Magallana gigas*	isolate 618	1205	1205	69%	0	100.00%	675
AF177226	*Crassostrea gigas*	*Magallana gigas*	mtDNA genome voucher	1205	1736	100%	0	100.00%	18224
MT219484	*Crassostrea gigas*	*Magallana gigas*	UHHCL21	1201	1201	69%	0	100.00%	651
MN862571	*Crassostrea gigas*	*Magallana gigas*	isolate EU9	1199	1199	69%	0	99.85%	655
MN862570	*Crassostrea gigas*	*Magallana gigas*	isolate EU8	1199	1199	69%	0	99.85%	655
Query sequence: 16S MW143047									
Accession	**Reported scientific name**	**Current scientific name**	**Isolate / Voucher**	**Max Score**	**Total Score**	**Query Cover**	**E-value**	**% Identity**	**Accession Lenght**
MN862573	*Crassostrea gigas*	*Magallana gigas*	isolate EU2	905	905	100%	0	100.00%	494
MF663018	*Crassostrea gigas*	*Magallana gigas*	isolate CGSC1b	905	905	100%	0	100.00%	540
MF663017	*Crassostrea gigas*	*Magallana gigas*	isolate CGSC1a	905	905	100%	0	100.00%	532
KJ855245	*Crassostrea gigas*	*Magallana gigas*	isolate WF34	905	905	100%	0	100.00%	18225
KJ855244	*Crassostrea gigas*	*Magallana gigas*	isolate YK05	905	905	100%	0	100.00%	18225
KJ855243	*Crassostrea gigas*	*Magallana gigas*	isolate YK01	905	905	100%	0	100.00%	18225
KJ855242	*Crassostrea gigas*	*Magallana gigas*	isolate JN14	905	905	100%	0	100.00%	18224
KJ855241	*Crassostrea gigas*	*Magallana gigas*	isolate CgJap23	905	905	100%	0	100.00%	18225
FJ478033	*Crassostrea gigas*	*Magallana gigas*	isolate CG38	905	905	100%	0	100.00%	511
EU672831	*Crassostrea gigas*	*Magallana gigas*	isolate ORCg-4	905	905	100%	0	100.00%	18225

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.19.449104; this version posted June 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

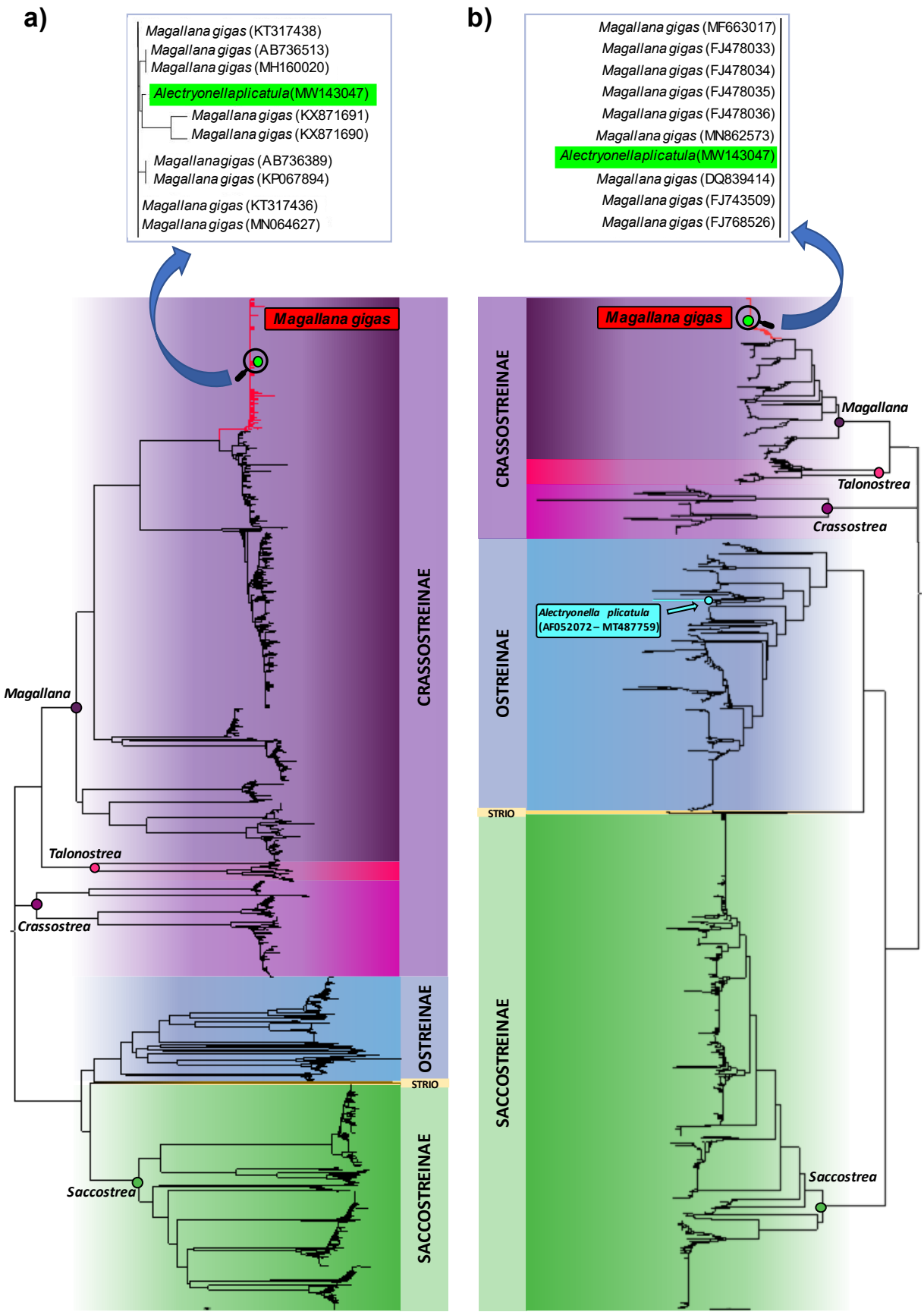

Figure 1. Neighbor-Joining trees based on 3639 *cox1* sequences (a) and 1839 16S sequences (b) available from public databases. In both trees MW143047 is nested within the clade formed by sequences of *Magallana gigas* within the Crassostreinae lineage. Instead, available 16S rRNA sequences of *Alectryonella plicatula* generated in previous studies cluster within the Ostreinae lineage. (STRIO:Striostreinae).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.19.449104; this version posted June 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

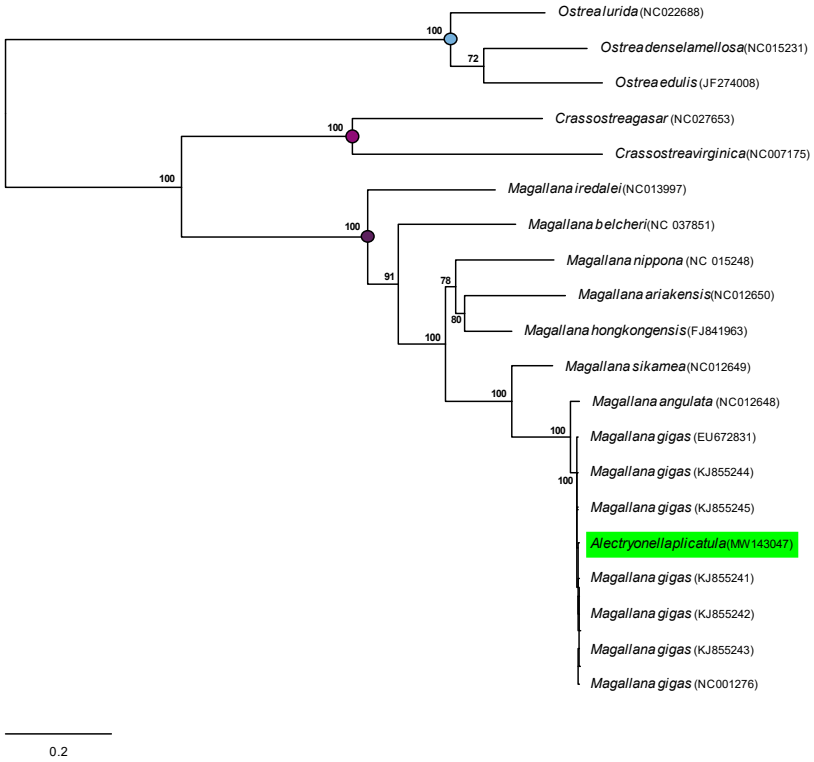

Figure 2. Maximum likelihood tree based on the concatenated sequences of 12 protein-coding from complete mitochondrial genomes of the same oyster taxa analyse d by Wang et al. (2021) plus six additional mitogenome sequences of *Magallana gigas*. The mitogenome MW143047 is nested with the clade formed by mitogenomes of *M. gigas*.

Discussion

Results of DNA barcoding, BLAST and phylogenetic analyses show that MW143047, attributed by Wang et (2021) to the fingerprint oyster *Alectryonella plicatula*, is identical to mitochondrial DNA sequences of the Pacific cupped oyster *M. gigas* (Table 1). The MW143047 sequences cluster within the clade of *M. gigas* both in the gene trees based on the barcoding markers *cox1* and 16S and in the ML mitogenome tree based on concatenated sequence of 12 PCGs (Fig 1 and 2). On the other hand, two mitochondrial 16S rRNA sequences of *A. plicatula* generated in previous studies (Jozefowicz & O'Foighil, 1998; Ardura et al., 2021), and available in Genbank under the accession numbers (AF052072 and MT487759), show a high genetic divergence (*p*-distance: 19 and 18% respectively) with MW143047. The most likely explanation for these results is that the original specimen used for mitogenome sequencing was misidentified and represents an individual of *M. gigas*.

The hypothesis of contamination by DNA of *M. gigas*, either prior to PCR amplification or as PCR product prior to sequencing, is unlikely. In these cases, often chimera sequence artefacts are observed (e.g., Sangster & Luksenburg, 2020), whereas all PCGs and rRNA genes of MW143047 are identical to sequences of *M. gigas* thus indicating that MW143047 is a *bona fide* mitogenome of

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.19.449104; this version posted June 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

M. gigas. Even less likely is the hypothesis of mitochondrial introgression of *M. gigas* in *A. plicatula* following hybridization. Indeed, while these two species might co-occur in the collection site of the original specimen used for sequencing (Shicheng Island, Dalian; China), their genetic divergence is very large (~19% at the 16S rRNA) as they belong to distinct evolutionary lineages within Ostreidae Rafinesque, 1815 (*A. plicatula* belongs to the Ostreinae lineage whereas *M. gigas* to the Crassostreinae lineage; e.g. O'Foighil & Taylor, 2000; Salvi et al., 2014; Crocetta et al., 2015; Salvi & Mariottini, 2017; Al-Kandari et al., 2021).

While *Magallana gigas* in *Alectryonella plicatula* are readily distinguishable using mitochondrial (Liu et al., 2011; Crocetta et al., 2015) or nuclear markers (O'Foighil & Taylor, 2000; Salvi et al., 2014; Mazón-Suástegui et al., 2016), morphological misidentification between the two might be easy as reported by Bishop et al (2017) due the extensive degree of phenotypic plasticity of oysters. This example highlights the common difficulties encountered for identifying oysters based on shell morphology alone, and provides one more demonstration that misidentification regards not only closely related species but also taxonomic ranks as high as subfamilies (discussed in Salvi & Mariottini, 2021; see Salvi et al. 2014 and Raith et al. 2016 for examples regarding the subfamilies Striostreinae Harry, 1985, Ostreinae Rafinesque, 1815 and Saccostreinae Salvi & Mariottini, 2016).

Previous studies on oyster systematics strongly advice that morphological identification of oysters should be validated by molecular data (e.g. Wang et al., 2004; Lam & Morton, 2006; Hamaguchi et al., 2107). This study also emphasizes that molecular data should be taxonomically validated by means of DNA barcoding and phylogenetic analyses. Taxonomic validation of mitogenomes is straightforward following the quality control guidelines of Botero-Castro et al. (2016) (see also Sangster & Luksenburg, 2020) and most of these recommendations can be applied also for an accurate identification of the sequences of single gene fragments. The publication of taxonomically misidentified sequences and mitogenomes can have profound implications if few sequences are available for the species so that misidentified sequence ends up as the reference for the species in public databases. In such cases misidentification errors can propagate in future studies that use the wrong reference-sequences in taxonomic and phylogenetic comparisons.

Disclosure statement

The authors report no conflict of interest.

ORCID

Daniele Salvi http://orcid.org/0000-0002-3804-2690
Emanuele Berrilli http://orcid.org/0000-0001-8081-8600
Matteo Garzia http://orcid.org/0000-0002-0918-9925
Paolo Mariottini http://orcid.org/0000-0003-1044-7108

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.19.449104; this version posted June 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Data availability statement

The data that support the findings of this study are openly available on GenBank at https://www.ncbi.nlm.nih.gov/nucleotide. Accession numbers of mitogenome sequences analysed are listed in Figure 2. Results of Blast and DNA-barcoding analyses are available from the authors upon request.

References

Al-Kandari M, Oliver PG, Salvi D., 2021. Molecular and morphological systematics of a new, reef forming, cupped oyster from the northern Arabian Gulf: *Talonostrea salpinx* new species. Zookeys, 1043: 1-20.

Ardura A, Fernandez S, Haguenauer A, Planes S, Garcia-Vazquez E. 2021. Ship-driven biopollution: how aliens transform the local ecosystem diversity in Pacifc islands. Marine Pollution Bulletin 166: 112251.

Bishop M, Brumbaugh R, Luckenbach M, Ruesink J. 2017. *Alectryonella plicatula*. The IUCN Red List of Threatened Species 2017: e.T200866A2683909. https://dx.doi.org/10.2305/IUCN.UK.2017-3.RLTS.T200866A2683909.en.

Bortolus A. (2008). Error cascades in the biological sciences: the unwanted consequences of using bad taxonomy in ecology. AMBIO: A journal of the human environment, 37(2), 114-118.

Botero-Castro F, Delsuc F, Douzery EJ. 2016. Thrice better than once: quality control guidelines to validate new mitogenomes. Mitochondrial DNA Part A. 27(1):449–454

Crocetta F, Mariottini P, Salvi D, Oliverio M. 2015. Does GenBank provide a reliable DNA barcode reference to identify small alien oysters invading the Mediterranean Sea?. Marine Biological Association of the United Kingdom. Journal of the Marine Biological Association of the United Kingdom, 95(1), 111.

Hamaguchi M, Manabe M, Kajihara N, Shimabukuro H, Yamada Y, Nishi E. (2017). DNA barcoding of flat oyster species reveals the presence of Ostrea stentina Payraudeau, 1826 (Bivalvia: Ostreidae) in Japan. Marine Biodiversity Records, 10(1), 1-10.

Hsiao ST, Chuang SC, Chen KS, Ho PH, Wu CL, Chen CA. 2016. DNA barcoding reveals that the common cupped oyster in Taiwan is the Portuguese oyster *Crassostrea angulata* (Ostreoida; Ostreidae), not *C. gigas*. Scientific Reports 6, 34057

Jin S, Kim KY, Kim MS, Park C. (2020). An assessment of the taxonomic reliability of DNA barcode sequences in publicly available databases. Algae, 35(3), 293-301.

Jozefowicz CJ, O'Foighil D. 1998. Phylogenetic analysis of southern hemisphere flat oysters based on partial mitochondrial 16S rDNA gene sequences. Molecular Phylogenetics and Evolution 10: 426–435

Kalyaanamoorthy S, Minh BQ, Wong TK, Von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature methods, 14(6), 587-589.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.19.449104; this version posted June 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

221 Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version
222 7.0 for bigger datasets. *Molecular biology and evolution*, *33*(7), 1870-1874.

223 Lam K, Morton B. 2006. Morphological and mitochondrial DNA analysis of the Indo-West Pacific
224 rock oysters (Ostreidae: *Saccostrea* species). Journal of Molluscan Studies 72: 235–245.

225 Liu JUN, Li QI, Kong L, Yu H, Zheng X. 2011. Identifying the true oysters (Bivalvia: Ostreidae)
226 with mitochondrial phylogeny and distance-based DNA barcoding. Molecular Ecology Resources,
227 11(5), 820-830.

228 Mazón-Suástegui JM, Fernández NT, Valencia I., Cruz-Hernández P, Latisnere-Barragán H.
229 (2016). 28S rDNA as an alternative marker for commercially important oyster identification. Food
230 control, 66, 205-214.

231 Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective
232 stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular biology and
233 evolution, 32(1), 268-274.

234 O'Foighil D, Taylor DJ. 2000. Evolution of parental care and ovulation behaviour in oysters.
235 Molecular Phylogenetics and Evolution 15: 301–313

236 Raith M, Zacherl DC, Pilgrim EM, Eernisse DJ. (2016). Phylogeny and species diversity of Gulf of
237 California oysters (Ostreidae) inferred from mitochondrial DNA. American Malacological Bulletin,
238 33(2), 263-283.

239 Ren J, Liu X, Jiang F, Guo X, Liu B. (2010). Unusual conservation of mitochondrial gene order in
240 Crassostrea oysters: evidence for recent speciation in Asia. BMC Evolutionary Biology, 10(1), 1-
241 14.

242 Salvi D, Macali A, Mariottini P. (2014). Molecular phylogenetics and systematics of the bivalve
243 family Ostreidae based on rRNA sequence-structure models and multilocus species tree. PLoS One,
244 9(9), e108696.

245 Salvi D, Mariottini P. (2017). Molecular taxonomy in 2D: a novel ITS2 rRNA sequence-structure
246 approach guides the description of the oysters' subfamily Saccostreinae and the genus *Magallana*
247 (Bivalvia: Ostreidae). Zoological Journal of the Linnean Society, 179(2), 263-276.

248 Salvi D, Berrilli E, D'Alessandro P, Biondi M. 2020. Sharpening the DNA barcoding tool through a
249 posteriori taxonomic validation: The case of *Longitarsus flea* beetles (Coleoptera: Chrysomelidae).
250 PloS one, 15(5), e0233573.

251 Salvi D, Mariottini P. (2021). Revision shock in Pacific oysters taxonomy: the genus Magallana
252 (formerly Crassostrea in part) is well-founded and necessary. Zoological Journal of the Linnean
253 Society, 192(1), 43-58.

254 Sangster G, Luksenburg JA 2020. The published complete mitochondrial genome of *Eptesicus*
255 *serotinus* is a chimera of *Vespertilio sinensis* and *Hypsugo alaschanicus* (Mammalia: Chiroptera).
256 Mitochondrial DNA Part B, 5:3, 2661-2664

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.19.449104; this version posted June 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

257 Wang H, Guo X, Zhang G, Zhang F. 2004. Classification of jinjiang oysters *Crassostrea rivularis*
258 (Gould, 1861) from China, based on morphology and phylogenetic analysis. Aquaculture 242: 137–
259 155

260 Wang Q, Liu H, Teng W, Yu Z, Liu X, Xie X, Yue C, Li D, Liang M, Li Q. 2021. Characterization
261 of the complete mitochondrial genome of *Alectryonella plicatula* (Bivalvia: Ostreidae).
262 Mitochondrial DNA Part B, 6(5), 1581-1582.