

Article

Mitochondrial Heteroplasmy and PCR Amplification Bias Lead to Wrong Species Delimitation with High Confidence in the South American and Antarctic Marine Bivalve *Aequiyoldia eightsii* Species Complex

Mariano Martínez ^{1,2,*}, Lars Harms ^{3,4} , Doris Abele ¹ and Christoph Held ¹

¹ Functional Ecology, Helmholtz Centre for Polar and Marine Research, Alfred Wegener Institute, Am Handelshafen 12, 27570 Bremerhaven, Germany

² Oceanografía y Ecología Marina, Instituto de Ecología y Ciencias Ambientales, Facultad de Ciencias, Universidad de la República, Iguá 4225, Montevideo 11400, Uruguay

³ Helmholtz Centre for Polar and Marine Research, Alfred Wegener Institute, Am Handelshafen 12, 27570 Bremerhaven, Germany

⁴ Helmholtz Institute for Functional Marine Biodiversity at the University of Oldenburg (HIFMB), Ammerländer Herrstrasse 231, 26129 Oldenburg, Germany

* Correspondence: mmartinez@fcien.edu.uy

Abstract: The species delimitation of the marine bivalve species complex *Aequiyoldia eightsii* in South America and Antarctica is complicated by mitochondrial heteroplasmy and amplification bias in molecular barcoding. In this study, we compare different data sources (mitochondrial cytochrome c oxidase subunit I (*COI*) sequences; nuclear and mitochondrial SNPs). Whilst all the data suggest that populations on either side of the Drake Passage belong to different species, the picture is less clear within Antarctic populations, which harbor three distinct mitochondrial lineages ($p\text{-dist} \approx 6\%$) that coexist in populations and in a subset of individuals with heteroplasmy. Standard barcoding procedures lead to amplification bias favoring either haplotype unpredictably and thus overestimate the species richness with high confidence. However, nuclear SNPs show no differentiation akin to the trans-Drake comparison, suggesting that the Antarctic populations represent a single species. Their distinct haplotypes likely evolved during periods of temporary allopatry, whereas recombination eroded similar differentiation patterns in the nuclear genome after secondary contact. Our study highlights the importance of using multiple data sources and careful quality control measures to avoid bias and increase the accuracy of molecular species delimitation. We recommend an active search for mitochondrial heteroplasmy and haplotype-specific primers for amplification in DNA-barcoding studies.

Keywords: mitochondrial heteroplasmy; amplification bias; mitochondrial DNA; DNA barcoding



Citation: Martínez, M.; Harms, L.; Abele, D.; Held, C. Mitochondrial Heteroplasmy and PCR Amplification Bias Lead to Wrong Species Delimitation with High Confidence in the South American and Antarctic Marine Bivalve *Aequiyoldia eightsii* Species Complex. *Genes* **2023**, *14*, 935. <https://doi.org/10.3390/genes14040935>

Academic Editor: Svenja Heesch

Received: 28 February 2023

Revised: 4 April 2023

Accepted: 5 April 2023

Published: 18 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The accurate delimitation and identification of species is pivotal to understanding the evolution of biodiversity and the response of communities to environmental change. DNA barcoding is a commonly used approach for assigning individuals to a known species based on standard reference DNA sequences (species identification) and for detecting still unknown species through deviations from known sequences (species delimitation). Ever since DNA barcoding was established [1], the sequencing of one or several standard DNA regions has been applied to many studies of phylogeny and phylogeography [2]. A fragment of the mitochondrial gene cytochrome c oxidase subunit I (*COI*) is one of the “gold standards” in the molecular barcoding of animals [3,4]. The utility of mitochondrial DNA (mtDNA) for species delimitation and identification is generally attributed to its high mutation rate, maternal inheritance, and lack of recombination [5].

Despite the widespread application of this convenient approach (as of February 2023, >12 million barcodes have been uploaded to the BOLD database [6]) and the invaluable progress in our understanding of global biodiversity patterns, the simplicity that prompted this technique's success is also what limits it. Assuming a strictly uniparental form of inheritance of mtDNA through the maternal germline can overlook phenomena such as the occurrence of pseudogenes or mitochondrial heteroplasmy, leading to incorrect phylogenetic and taxonomic inferences. The risk of inadvertently using nuclear mitochondrial pseudogenes (Numts) in barcoding analyses has been considered to some degree in the barcoding literature [7,8]. However, the existence of different mitochondrial genomes in a single organism (heteroplasmy) has scarcely been suggested as a source of bias [9,10]. To the best of our knowledge, there are no concrete examples in the literature in which this peculiar feature of the mitochondrial genome has been identified as a source of misinterpretation of patterns of genetic divergence. There could be many reasons for this, but the exclusion of sequencing results characterized by the superposition of several templates in heteroplasmic organisms may be a contributing factor.

Heteroplasmy has been found in a wide range of taxa ranging from mollusks and arthropods [10–12] to vertebrates including fish, birds, and mammals [13–15]. The best-studied heteroplasmic system in metazoans is the Doubly Uniparental Inheritance (DUI) of mitochondria, an evolutionarily stable mechanism distributed across over 100 bivalve species [16–18]. In bivalves with DUI, females normally inherit mitochondria from their mothers (F-type), whilst males inherit mitochondria from both their mother and father (F-type and M-type). In males, the F-type is dominant in somatic tissues, whereas the M-type is prominent in the gonads but can also occur in small amounts in the somatic tissue [19]. Sex-associated mtDNAs can be extremely divergent, with up to 50% intraspecific DNA divergence [20], which is well above the divergence level typically observed among bivalve species [21].

DUI is widespread within bivalve subclasses, including among Protobranchia [22], the most basal extant bivalves with a poorly resolved phylogeny [23]. Protobranchs are of particular interest because of their high abundance in the deep oceans and important role in surface sediment carbon turnover [24]. The high-latitude marine protobranch *A. eightsii* (previously of the genus *Yoldia*) is widely distributed in the Southern Ocean, inhabiting the soft-substratum ecosystems of South America, the Antarctic Peninsula, and several Subantarctic islands. This wide distribution range on both sides of the Drake Passage has given rise to several revisions of its taxonomic status [25]. Recently, Muñoz-Ramírez et al. [26] highlighted the role of the Antarctic circumpolar current as a biogeographic barrier between the Antarctic Peninsula and South America, confirming the existence of two different species across the Drake Passage, which, according to Martínez et al. [27], have different levels of susceptibility to two major climate change stressors, namely, temperature shifts and hypoxia. González-Wevar et al. [28] suggested the presence of several cryptic species comprising two lineages along the Antarctic Peninsula (5.78% COI p-distance), a third lineage in South America (6.5–7.5% COI and 1.2% ITS p-distances) with respect to the Antarctic peninsula, and two additional mitochondrial lineages on Kerguelen Island and the Falkland/Malvinas Islands. However, the scarcity of available nuclear information has hampered formal taxonomic delimitations thus far; hence, *A. eightsii* is still considered a single species.

The advances in speed and accuracy in tandem with reduced costs of next-generation sequencing (NGS) technologies have facilitated the generation of population-scale genomic surveys. Transcriptome sequencing (RNA-seq) is particularly useful for studying non-model organisms lacking reference genomes [29]. Recently, the number of studies using single-nucleotide polymorphisms (SNPs) from RNA-seq data has increased significantly, thus establishing transcriptomics as an alternative data source for population genomic studies [30–32]. The key analytical advantage of gaining population genomic insights from expressed sequences is the reduction in complexity afforded by limiting the sequence data to a smaller fraction devoid of repetitive and heterochromatic sequences [33].

Based on the hypothesis of heteroplasmy being a potential confounding factor in population genetics and phylogenetic analyses, this study evaluates patterns of genetic divergence between South American and Antarctic *A. eightsii* populations. Our analysis compares the results of a classic barcoding approach based on selected mitochondrial and nuclear gene fragments with SNPs derived from the deep sequencing of expressed genes, with a particular focus on how instances of heteroplasmy, if any, may affect population genetic and phylogenetic inferences.

2. Materials and Methods

2.1. Animal Collection

A. eightsii samples were collected in South America (Magellanic region) and Antarctic Peninsula in the austral summer season of October 2017 until February 2018. Bivalves from the Magellanic region ($n = 100$) were collected by SCUBA divers in the shallow subtidal zone near Punta Arenas (PA; Chile, 53°37'52" S; 70°56'54" W) on a single day in October 2017. Antarctic bivalves ($n = 104$) were collected between January and February 2018 from three sites in Potter Cove (PC), an 8 km² glacial fjord on King George Island, South Shetlands (62°14'11" S; 58°40'14" W–62°13'32" S; 58°38'31" W–62°13'35" S; 58°40'58" W), in shallow waters between 6 and 20 m using a Van Veen grab. The maximum distance between the three collection sites in PC was 1 km, and the sites were sampled within approximately one week. An additional subset of samples ($n = 34$) collected in Hangar Cove in February 2007 (HC; near Rothera station, Adelaide Island) and Potter Cove in January 2016 and conserved at −20 °C was provided by colleagues from the British Antarctic Survey and the University of Cordoba (Argentina), respectively. These samples were exclusively used for the mitochondrial and nuclear loci analyses. In total, we obtained 238 individual organisms.

During each sampling period, bivalves were transported to the local research facility in insulated containers containing seawater and sediment from the sampling site and were dissected under a stereomicroscope after recording their individual weights, shell lengths, and widths. For each individual, three tissue types were conserved: foot and mantle tissue samples were conserved separately in RNAlater (SIGMA; St. Louis, MO, USA) and stored at −80 °C, and the rest of the soft body tissue was stored in Ethanol 95% at 4 °C until further analysis. In total, 168 bivalves were included in the analysis. All of them were utilized for the mitochondrial and nuclear marker analyses, and a subset of 70 individuals were analyzed to obtain SNPs via RNA-seq data.

2.2. Genomic DNA Extraction and Amplification

The molecular markers included one mitochondrial gene (cytochrome c oxidase subunit I, *COI*) amplified using Folmer primers and haplotype-specific primers (see Section 2.5 below), one nuclear protein-encoding gene (*histone H3*), and one nuclear ribosomal gene (18S ribosomal RNA, *18S*). Total genomic DNA was extracted from muscle tissue with QIAamp DNA Mini Kit (QIAGEN; Hilden, Germany) following the manufacturer's protocol (regarding DNA purification from tissues). *COI* amplification and sequencing was performed for 155 individuals ($n_{PA} = 50$, $n_{PC} = 82$, and $n_{HC} = 24$); *histone H3* and *18S* fragments were amplified and sequenced for 72 animals ($n_{PA} = 24$, $n_{PC} = 24$, and $n_{HC} = 24$). DNA purity and concentration were assessed using a Nanodrop Spectrophotometer® ND-1000 (NanoDrop Technologies; Wilmington, DE, USA). Purified genomic DNA was diluted to 10 ng/μL and used as template for Polymerase Chain Reaction (PCR) amplification. All PCRs were carried out in 25 μL reaction volumes comprising 5.0 μL 5X Colorless GoTaq® Flexi Buffer (Promega Corp; Madison, WI, USA), 2 μL of MgCl₂ (25 mM), 2.5 μL of dNTP mix (2 mM), 0.25 μL of Betain (5M), 0.125 of each forward and reverse primer (100 μM), 0.15 μL of GoTaq® G2 Flexi DNA Polymerase (5 U/μL, PROMEGA Corp; Madison, WI, USA), and 3 μL of DNA extract (10 ng/μL). A fragment of 640 bp corresponding to *COI* was amplified using Folmer primers [3] through a PCR program consisting of the following steps: 2 min at 95 °C followed by 35 cycles at 95 °C for 20 s, 46 °C for 20 s, 72 °C

for 40 s, and a final extension of 8 min at 72 °C. For the amplification of *histone H3* and *18S* fragments (257 and 752 bp, respectively), new primers were designed based on sequences of species from the same family published in [23] (*histone H3*: forward primer: 5'-GAA AAT CTA CCG GTG GCA AG-3', reverse primer: 5'-GTG TCC TCG AAC AAA CCA AC-3'; *18S*: forward primer: 5'-AAG TAC AGA CTC TCA GTA CCG-3', reverse primer: 5'-GAA GGC CAA CAA AAT AGA ACC-3') and the PCR program used for both fragments consisted of the following steps: 2 min at 95 °C, 35 cycles at 95 °C for 20 s, 56 °C for 20 s, 72 °C for 40 s, and a final step of 8 min at 72 °C. After the amplification, an initial quality assessment of the PCR was conducted by placing the product on 2% agarose gel. All PCR products were sequenced for both strands by Eurofins Genomics GmbH (Ebersberg, Germany) using the same primers that were utilized during amplification. *COI*, *histone H3*, and *18S* sequences were aligned using the ClustalW algorithm of CodonCode Aligner program (version 5.1.5; CodonCode Corporation, Dedham, MA, USA), and all chromatograms were inspected visually for sequencing mistakes. For *COI*, a haplotype network was constructed using the Neighbor-Joining algorithm and implementing the Hasegawa–Kishino–Yano (HKY) model of substitution in Geneious using default parameters (version 8.1.9, Biomatters Ltd., Auckland, New Zealand) and Haplotype Viewer (Center of Integrative Bioinformatics Vienna, <http://www.cibiv.at> (accessed on 6 April 2022)).

2.3. RNA Extraction and cDNA Library Preparation

A subset of 70 individual samples out of the total included in mitochondrial and nuclear loci analysis were analyzed to obtain SNPs via RNA-seq data. A total of 70 transcriptomic libraries (one library per individual) were created in this study for animals from Punta Arenas and Potter Cove ($n_{PA} = 29$, $n_{PC} = 41$)

Samples of mantle tissue (5–30 mg) were homogenized in Trizol reagent (SIGMA; St. Louis, MO, USA) using a Precellys homogenizer (Precellys24, Bertin Technologies, Paris, France). Total RNA was isolated from each sample using the Direct-zol™ RNA MiniPrep Kit (ZYMO Research Corp., Irvine, CA, USA) according to the manufacturer's instructions. The concentration and quality of the RNA were determined using a Nanodrop Spectrophotometer ND-1000 (NanoDrop Technologies, Wilmington, DE, USA) and LabChip® GX Touch (PerkinElmer; Waltham, MA, USA). Libraries were prepared using the Illumina TruSeq® Stranded mRNA Sample Preparation Kit starting from 1 µg of total RNA. The Poly-A-containing mRNA molecules were purified using poly-T oligo-attached magnetic beads. Subsequently, the mRNA was fragmented using divalent cations under elevated temperature and copied into the first strand of cDNA using reverse transcriptase and random primers. This process was followed by second-strand cDNA synthesis using DNA Polymerase I, RNase H, and dUTP instead of dTTP to achieve strand specificity and remove the RNA template. Following adapter ligation, products were enriched by PCR and purified to create a cDNA library. Libraries were validated and quantified using a LabChip® GX Touch (PerkinElmer, Waltham, MA, USA). All the samples were pooled and cleaned using magnetic beads to remove the remaining primer content. Final cDNA concentration was measured using LabChip® GX Touch (PerkinElmer, Waltham, MA, USA). The pool of samples was sequenced at the Alfred Wegener Institute on an Illumina NextSeq 500 sequencer using the NextSeq High Output Kit v2 (150 cycles) with a paired-end protocol.

2.4. De Novo Assembly and SNP Analysis of RNA-Seq Data

Raw reads were quality-controlled by FastQC v. 0.11.7 (Babraham Institute, Cambridge, UK). Adapter sequences were removed using bbdutk.sh from the BBtools suite (version 36.38) [34] with the following parameters: ktrim = r, k = 23, mink = 11, hdist = 1, tpe, and tbo. Sort-MeRNA version 2.1 was used to search for rRNA sequences in the remaining sequences [35], which were removed before further processing. To filter the sequences for the common Illumina spikein PhiX, bbdutk.sh was used with a kmer size of 31 and a hdist of 1. A final quality-trimming procedure was performed with bbdutk.sh using Q10 as minimum quality and 36 bases as the minimum length. All obtained sequences (70 libraries) were normal-

ized using `bbnorm.sh` [34] with an average depth of $100\times$ and a minimum depth of $5\times$; subsequently, they were de novo assembled using the Trinity genome-independent transcriptome assembler version 2.8.4 [36] with a minimum transcript length of 300 bases and the option for strand specificity selected (`-SS_lib_type RF`). To remove duplicate sequences from the assembly, `dedupe.sh` [34] was used with the following parameters: `minidentity = 98`, `arc = t`, `am = t`, and `ac = t`. The read representation and strand specificity of the assembly were assessed using the software Bowtie2 v.2.3.4.1, and completeness was evaluated using the package BUSCO v 4.1.4 (Benchmarking Universal Single-Copy Orthologs) and the orthologs of the public database “metazoa_odb10”.

For the SNP analysis, quality-filtered paired-end reads were aligned to the de novo transcriptome using `bowtie2` v2.3.4.1 [37]. Alignments in SAM format (Sequence Alignment Map) were compressed and indexed with `SAMtools` v1.8 [38]. Genotype likelihoods were computed using `mpileup` from `SAMtools` and variant calling was performed using `BCFtools`. In the first filtering procedure, we excluded all variants with a Phred quality score below 30. Since our aim was to analyze SNPs, we selected only those SNPs from the variant calling that were present in at least one individual. A primary dataset of 1,308,131 SNPs in 70 individuals was filtered using `VCFtools` v0.1.16 [39] and by applying an iterative filtering strategy between loci and individuals with a progressive increase in cut-off values [40]. An initial filtering procedure for loci quality was performed, in which variants successfully genotyped in 50% of the individuals were kept (max-missing 0.5) and a minor allele count (MAC) of 3 was employed. Subsequently, individuals with more than 37% missing data were excluded from the analysis. A second filtering procedure for loci quality implied a maximum missing value of 0.95, a Minor Allele Frequency (MAF) of 0.05, and a minimum number of reads (minDP) equal to 10. Additionally, variants with more than one allele were discarded (max-alleles 2). Nuclear and mitochondrial SNPs were separated by blasting the sequences containing SNPs against a mitochondrial database obtained from the UniProt Swiss-Prot database. A total of 113,515 SNPs in a final number of 54 individuals were retained. Of these, 4714 were identified as mitochondrial SNPs, while nuclear SNPs comprised all SNPs in sequences with no hits in the mitochondrial database (108,801). Mitochondrial SNPs included in the analysis (392) were those contained in a single long transcript (ca 18 kbp), representing most of the mitochondrial genome (annotated using MITOS following the approach reported by Bernt et al. [41] shown in Supplementary Table S1). To assess genetic differentiation between South American and Antarctic animals, but also between animals of different mitochondrial genotypes (mitotypes, see Section 2.5), Principal Component Analyses (PCAs), Analysis of Molecular Variance (AMOVA), and pairwise G_{ST}' estimations were carried out separately for nuclear and mitochondrial SNPs with all the individuals that remained after the filtering process. G_{ST}' is a corrected version of the coefficient of genetic differentiation G_{ST} , which is a quotient of heterozygosity estimates obtained from a subpopulation and the whole population. A corrected value $G_{ST}' = G_{ST}/G_{STmax}$ was computed because G_{ST} will never reach the theoretical maximum of 1 under ordinary circumstances [42]. Additional PCAs excluding the South American samples were performed to enhance differences between Antarctic mitotypes resulting from our analysis. AMOVA was conducted with respect to three strata, namely, location (South America and Antarctica), mitotype, and individuals, and significance was tested by randomly permutating ($n = 1000$) the sample matrices as described in Excoffier et al. [43]. All the analyses were performed in R v3.6 [44] using the packages `vcfR` [45], `adegenet` [46], `poppr` v2.8.3 [47], `ade4` [48], and `pegas` [49].

2.5. Coexistence of Mitochondrial Variants in a Single Individual

Observations of raw SNP data (reads aligned to de novo transcriptome, see Section 2.4) suggested the coexistence of more than one mitochondrial haplotype (of those previously defined based on *COI* data; see Section 3.1) in a subset of individuals from Potter Cove (Antarctica). These specimens carry the haplotype h1 or h2 (never both) together with haplotype h3. To verify this finding, haplotype-specific primers were designed to selectively

amplify the existing mitochondrial variants in individual samples and avoid artifacts caused by competitive differences of either mitochondrial variant in a competitive PCR using a single Folmer primer pair for both amplicons. Primer design was performed using a long transcript (ca. 18 kbp) of the de novo transcriptome representing most of the mitochondrial genome, wherein the primers were positioned in regions that were variable among haplotypes (3 bp difference). Thus, two forward primers were designed to amplify haplotype h3 (5'-AAT GTT AAT TTG TTC CAT GAG G-3'; Pb) and haplotypes h1, h2, and h4 (5'-AAT GTT AAT TTG TTC TAT GGC G-3'; Pa) and used in combination with one common reverse primer (5'-AGA AAA TAC AGC CCC CAT TC-3') designed for both amplifications. The target amplicon (714 bp) was intentionally overlapped (by ca. 300 bp) with the *COI* sequence amplified with Folmer primers in order to track the haplotype identification based on *COI*. A subset of samples of the four haplotypes from every location ($n_{h1} = 31$, $n_{h2} = 27$, $n_{h3} = 5$, and $n_{h4} = 16$) were amplified with both primer combinations. PCR conditions for both primer combinations were the same as for *COI* with Folmer primers, with the exception of the annealing temperature (56 °C). The size of PCR products was determined on 2% agarose gel, and a representative subset of samples of the mitochondrial haplotypes successfully amplified for one or two variants ($n = 32$) was sequenced on both strands by Eurofins Genomics GmbH (Ebersberg, Germany). Finally, samples were aligned using the ClustalW algorithm of CodonCode Aligner program (version 5.1.5; CodonCode Corporation, Dedham, MA, USA). This analysis allowed us to group the animals carrying a single or two combined haplotypes in different genetic clusters (mitotypes), which was further considered for SNPs.

3. Results

3.1. Mitochondrial and Nuclear Sequence Analysis

There were no differences between the *histone H3* sequences, and only one polymorphic site was found in the *18S* fragment, which did not correspond to any geographic pattern in our sampling nor any of the clades suggested by the mitochondrial or nuclear SNP data (see Section 3.3). In most of the cases, the Sanger sequencing of the *COI* amplified using Folmer primers yielded clear sequencing results. In some cases ($n = 13$), however, ambiguous base calls that suggested the possibility of the co-existence of several *COI* alleles in the template (i.e., heteroplasmy; see 3.2 below) were observed.

The results of the *COI* sequencing with the Folmer primers show a clear differentiation between Antarctic and South American individuals (h4) but also a division between two major Antarctic clades (h1 + h2 vs. h3; Figure 1). The overall haplotype network includes 15 distinct haplotypes, 4 in South America (PA) and 9 in Antarctica, of which 7 were exclusive to PC or HC, whereas 2 were present at both Antarctic locations. The most frequent *COI* haplotypes (>2 individuals) are labeled in Figure 1 (h1, h2, h3, and h4), and their genetic distances (p-distances) are shown in Supplementary Table S2. The high p-distances between the South American haplotype h4 and the Antarctic ones (6.6–8.6%) and the distances between the rare h3 and the more frequent Antarctic h1 and h2 haplotypes (5.9–6.1%) were remarkable. These differences represent four amino acid substitutions between the Antarctic haplotypes h1 and h2 and the South American h4 and five between the Antarctic haplotype h3 and h4. No amino acid substitutions were found between the Antarctic haplotypes h1 and h2, whereas five substitutions were found between these two haplotypes and h3. No stop codons or indels were found in the *COI* sequences of any haplotype or location.

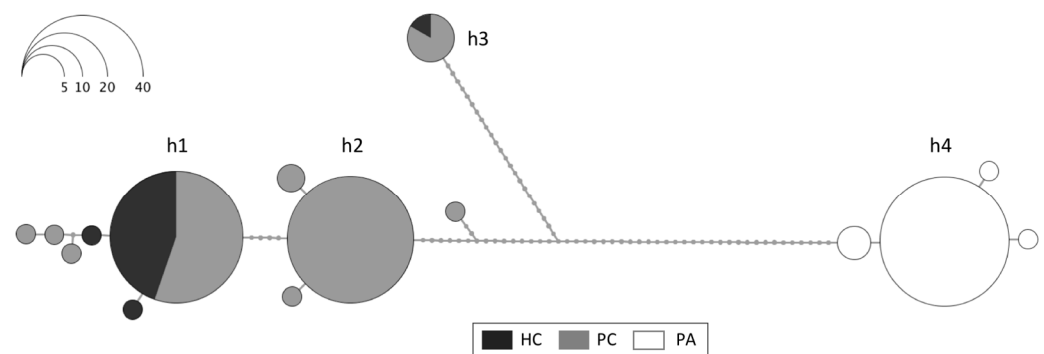


Figure 1. Haplotype network based on *COI* sequences of *Aequioidia bivalves* ($n = 155$) amplified with Folmer primers from Hanger Cove (HC), Potter Cove (PC), and Punta Arenas (PA). Each haplotype is represented by a circle proportional to its frequency (scale top left), and the most frequent haplotypes (>2 individuals) are labeled as h1, h2, h3, and h4.

3.2. Coexistence of Mitochondrial Variants in a Single Organism

The co-existence of two haplotypes in a single individual was detected in several Antarctic individuals from PC (19 out of 47) and HC (1 out of 16) (h1, h2, and h3) but not in South American organisms (0 out of 16) (h4). In all but one individual that had been identified as carrying the haplotype “h3-only” with Folmer primers (see Section 3.1), the haplotypes h1 or h2 were also found once haplotype-specific primers were used. A single bivalve from HC, in which variants h1 or h2 could not be amplified, was interpreted as possessing haplotype h3 only (h3 homozygote). The reverse case, in which the Folmer primer only detected h1 or h2 but left h3 undetected, was also observed. All South American specimens (haplotype h4) and the majority of Antarctic animals analyzed qualified as homoplasmic for h1 or h2 with Folmer primers and with the newly designed haplotype specific primers (Figure 2). These findings allowed us to group the animals in six different genetic clusters based on their mitochondrial genotypes (mitotypes): animals carrying only one haplotype (h1, h2, h3, and h4) are henceforth named h1h1, h2h2, h3h3, and h4h4, respectively, and heteroplasmic animals for haplotype h1 or h2 together with the haplotype h3 are henceforth called h1h3 and h2h3, respectively. Mitotypes frequencies at each location are shown in Figure 2. This mitotype-based grouping was considered in the SNP analysis.

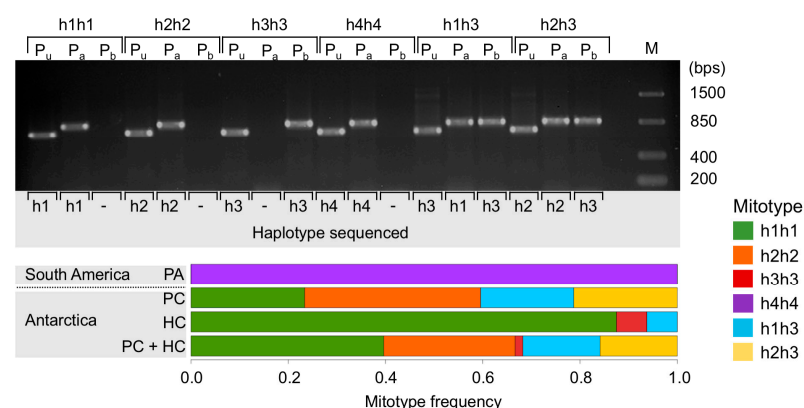


Figure 2. Above: Gel electrophoresis (2% agarose) of PCR-amplified products from individual samples of the six mitotypes (h1h1, h2h2, h3h3, h4h4, h1h3, and h2h3) using Folmer primers for *COI* (Pu), and the two haplotype-specific primer combinations designed in this study (Pa: h1-, h2-, and h4-specific; Pb: h3-specific). Sanger sequencing results (haplotypes) of each PCR product are shown in the bottom of the gel. Below: mitotype frequencies of bivalves from Punta Arenas (PA, $n = 16$), Potter Cove (PC, $n = 47$), Hanger Cove (HC, $n = 16$), and both Antarctic locations together (PC + HC, $n = 63$).

3.3. De Novo Assembly and SNP Analysis

The Illumina sequencing resulted in 454.8 million reads (average reads per library: 6.5 million), of which 388.4 million reads were used for the de novo transcriptome assembly after being filtered for quality. A total of 389,929 transcripts were assembled, with an average length of 764 bp. The BUSCO assessment of the completeness of the transcriptome showed very high quality, for which 98.7% (Complete BUSCOs: 96.6%, Fragmented BUSCOs: 2.1%) of the orthologs of the public database “metazoa_odb10” were present in the assembly.

The results of the Principal Component Analysis (PCA) for nuclear and mitochondrial SNP data are shown in Figure 3. The graphic representation of the nuclear and mitochondrial PCA includes the assignment of mitotypes to each individual plotted. Both the nuclear and mitochondrial SNP compositions clearly separate the South American animals (h4h4) from the Antarctic ones along PC1, explaining 58.6% and 88.6% of the variance, respectively. In contrast, the nuclear SNPs of the Antarctic animals were not congruent with the mitotype grouping; instead, three groups composed of animals of the four Antarctic mitotypes could be discerned. These groups, which were separated by a low percentage of variance (PC1: 5.6%, PC2: 4.8%), partially matched the sampled sites within Potter Cove, indicating close kinship due to proximity. The PCA of the mitochondrial data resulted in two clusters, one containing animals of the mitotypes h1h1 and h1h3 and the other animals corresponding to h2h2 and h2h3, with no signal reflecting the three sampling sites in Potter Cove. The removal of South American animals corresponding to h4h4 from the analysis resulted in a clearer differentiation of the Antarctic mitotypes h1h3 and h2h3 vs. h1h1 and h2h2 (PC2).

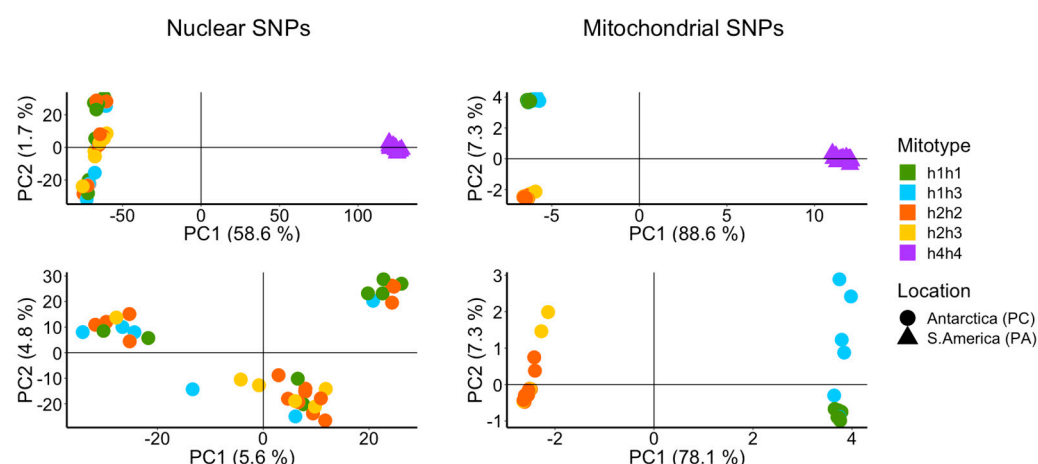


Figure 3. Principal component analysis based on 108,801 nuclear and 392 mitochondrial SNPs of Antarctic (circles) and South American (triangles) populations of *A. eightsi*. In panels below, only Antarctic specimens were included in the analysis. Nuclear and mitochondrial SNPs identify the samples from South America as a distinct group. Mitochondrial SNPs identify two groups among Antarctic samples characterized by haplotypes h1 or h2, respectively, either in homoplasmic condition or in heteroplasmic condition together with h3. Nuclear SNPs sort Antarctic samples into three groups that are not congruent with the mitochondrial conditions of the samples. Mitochondrial genotypes (Mitotypes) of the organisms are indicated with colors in both PCAs based on nuclear and mitochondrial SNPs.

The AMOVA results confirmed the clear patterns observed in the PCAs. For mitochondrial SNPs, all three levels of variance partitioning were highly significant ($p < 0.001$), with most of the variation occurring between locations (South America and Antarctica, 89.6%) and a minor proportion occurring between mitotypes within locations (7.8%) and among individuals (2.6%). The AMOVA performed on the nuclear SNPs presented significant differences between locations (67.7%) and among individuals (31.7%) but not between mitotypes within locations. This information is summarized in Table 1.

Table 1. Coefficients of the Analysis of Molecular Variance (AMOVA) of nuclear (Nuc) and mitochondrial (Mit) SNPs considering three strata (location, mitotype, and individual).

Source of Variation	df		Sum of Squares		Percentage of Variation		p Value	
	Nuc.	Mit.	Nuc.	Mit.	Nuc.	Mit.	Nuc.	Mit.
Between location	1	1	4.33	10.33	67.73	89.58	0.001	0.001
Between mitotype, within location	3	3	0.09	0.30	0.57	7.82	0.442	0.001
Within individual	49	49	0.08	0.012	31.70	2.60	0.001	0.001

df: degree of freedom.

Both the nuclear and mitochondrial pairwise genetic differentiation estimates (G_{ST}') showed a high density of highly differentiated SNPs ($G_{ST}' = 1$; i.e., present in all animals of one group and in none of the other) between animals from the two continents (Figure 4a,b), thus supporting the AMOVA, PCA, and *COI* results. An important difference between nuclear and mitochondrial G_{ST}' estimations emerged in the form of a high density of nuclear SNPs with values close to zero in comparison to the mitochondrial data. Nuclear and mitochondrial pairwise G_{ST}' estimates between Antarctic mitotypes are shown in Figure 4c,d. The nuclear pairwise G_{ST}' comparisons resulted in similar density distributions, with most of the values being <0.2 and the means ranging from 0.05 to 0.09, indicating low genetic differentiation. Instead, every mitochondrial G_{ST}' estimate comparing animals carrying the haplotype h1 and animals with h2 resulted in a high density of highly differentiated SNPs ($G_{ST}' = 1$). G_{ST}' estimations comparing h1h1 vs. h1h3 and h2h2 vs. h2h3 showed a low level of genetic differentiation (Figure 4d).

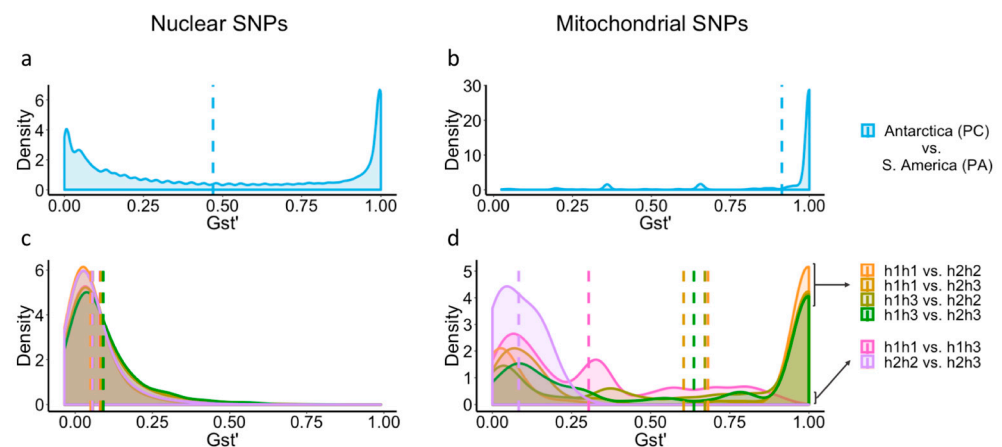


Figure 4. Density distributions of G_{ST}' values estimated by pairwise comparison between South American and Antarctic populations of *A. eightsii* based on nuclear and mitochondrial SNPs (a,b) and between Antarctic mitotypes based on nuclear and mitochondrial SNPs (c,d). Dashed lines represent the mean G_{ST}' of the correspondingly colored distribution. Note that to ease visualization, comparisons between continents do not involve comparisons within Antarctic mitotypes, and comparisons between mitotypes do not include the mitotype from South America (h4h4). A combined graphic representation would illustrate a bimodal distribution for both data sets (mitochondrial and nuclear SNPs), with most of the values close to zero for every comparison within Antarctic mitotypes and a high proportion of values close to one for every comparison between Antarctic and South American groups.

4. Discussion

This study calls into question the reliability of a species identification and delimitation system based on a few nuclear or mitochondrial markers such as the widely used cytochrome c oxidase subunit I (*COI*). We present a case study in a non-model marine bivalve, in which the standard barcoding approach using Folmer primers is systematically biased by the occurrence of several distinct mitochondrial genomes in the population,

including heteroplasmy, which, in the absence of independent data (e.g., nuclear SNPs), would lead to an erroneous interpretation of the number of species and their distribution patterns (Figure 5). Since this systematic bias is not caused by a paucity of data, it is not averted by increasing the number of data (i.e., the number of individuals sequenced) either; instead, it might converge, as in our case and as presented in previous studies [28], on the wrong solution with high confidence. This will be especially common in non-model species whose genomic features are mostly unknown and for which inferences and indiscriminate generalization from model organisms are commonly used procedures [50].

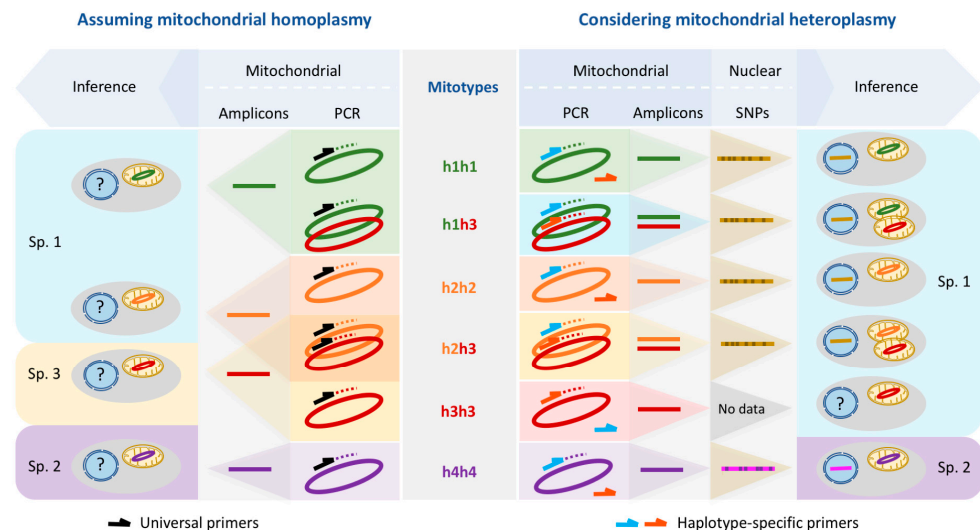


Figure 5. The amplification bias of Folmer primers and a standard barcoding analysis assuming strictly clonal inheritance of mitochondria leads to an overestimate of inferred species numbers (left half) in comparison to haplotype-specific primers and the inclusion of nuclear SNP data (right half). Note that amplifying h2h3 individuals using Folmer primers leads to the amplification of either h2 or h3, but never both. Although no nuclear SNP data are available for the single h3h3 individual, we include it in species 1 here on the grounds that h3 mitochondria are associated with nuclear SNP patterns that characterize species 1 (see h1h3 and h2h3).

4.1. Incongruence of Methods

4.1.1. Comparison across the Drake Passage

Both mitochondrial and nuclear SNP analysis showed significant differences between Antarctic and South American bivalves, thus confirming the clear pattern observed in the PCAs and in the GST' estimations. The bimodal distribution of pairwise differences in both data sources demonstrate that South American and Antarctic populations are genetically much more similar within the limits of either continental waters but highly distinct when compared between them. When viewed in conjunction with the marked differences in shell size (see Supplementary Figure S1), the most likely conclusion is that populations on either side of the Drake Passage are genetically isolated and represent distinct species. Considering that the locus typicus (New South Shetland) for *A. eightsii* (Jay, 1839) is close to our sampling location (Potter Cove), it is reasonable to assume that all our Southern Ocean samples belong to *A. eightsii* (Jay, 1839) sensu strictu, whilst our South American samples (haplotype h4) belong to an as yet undescribed species that has been confused with it. However, a full taxonomical revision including the formal description of the new species of *Aequiyoldia* is beyond the scope of this paper.

4.1.2. Comparison within the Southern Ocean

Whilst the extrapolation from one mitochondrial marker gene to the assumed congruent differentiation of the nuclear genome that the standard barcoding procedure implies

seems justified in the comparison across the Drake Passage (Figure 4), the same is not true for emerging patterns of differentiation in the Antarctic.

In the Southern Ocean, the most notable patterns of genetic differentiation in *A. eightsii* are strongly dependent on the source of data used. Although a differentiation pattern of a magnitude similar to that of the species-level trans-Drake comparison is also present within the Southern Ocean in the mitochondrial *COI* sequence barcoding data (Figure 1; haplotypes (h1 + h2) vs. h3), the nuclear SNP data show no indication of such differentiation but show a unimodal distribution that is characteristic of a single genetic entity (Figure 4c) for all specimens regardless of their mitochondrial haplotypes (h1, h2, or h3). Furthermore, the mitochondrial *COI* barcoding sequences (derived from a 600 bp fragment of the mitochondrial *COI* gene) and the mitochondrial SNPs (derived from the 18 kb transcript representing the almost-complete mitochondrial genome) suggest a mutually incompatible differentiation pattern. The *COI* barcode sequences suggest a structure corresponding to (h1 + h2) vs. h3, whereas for the mitochondrial SNPs the most apparent division is orthogonal to the *COI* sequence data, namely, (h1h1 + h1h3) vs. (h2h2 + h2h3), possibly h3h3 by itself; thus, individuals are most strongly characterized by possessing either h1 or h2 with varying proportions of h3 admixed with either of them. This finding is not only different but incongruent and cannot easily be reconciled by assuming a different degree of taxonomic resolution in the various data sources.

4.1.3. Taxonomic Consequences

The classical molecular barcoding approach based on the amplification and sequencing of a fragment of the mitochondrial *COI* gene turned out to be a good first indicator of major genetic diversity in the trans-Drake comparison. Instead, the heteroplasmic mitochondrial condition observed in some Antarctic specimens significantly interfered with its taxonomic classification when addressed by traditional molecular approaches. Considering the magnitude of the differences found by González-Wevar et al. [28] based on *COI*, it seems likely that they, too, had detected the division between h3 and (h1 + h2) and hence suggested that these two lineages were separate species, a conclusion that is not supported by our more comprehensive dataset. Our mitochondrial and nuclear SNP dataset and the coexistence of haplotype h3 with either haplotype h1 or h2 in some organisms exclude cryptic speciation between these two groups as an explanation for the observed patterns.

The widespread occurrence of mitochondrial heteroplasmy linked to Doubly Uniparental Inheritance (DUI) in bivalves and the evidence of mitochondrial heteroplasmy in a wide variety of taxa from crustaceans to vertebrates [12,15] highlight the importance of mitochondrial heteroplasmy as a general, previously underestimated shortcoming of molecular barcoding methodology. There is a consensus concerning the need for multiple loci (mitochondrial and nuclear) to perform a correct taxonomic classification, and the use of advanced sequencing techniques is intended to gain space in molecular systematics to amply address this need. However, the promise of this new sequencing era has not yet fully materialized, especially not for non-model species, demonstrating a growing taxonomic bias in favor of a small minority of genetic model species [51]. DNA barcoding using mitochondrial markers is still the dominant method with respect to non-model species and will probably remain so for a while. *A. eightsii* serves as a model case demonstrating that if heteroplasmy occurs, it can lead to false taxonomic inferences with an inflated degree of confidence. Correct taxonomic delimitation and identification is essential for the monitoring and conservation of biodiversity, especially in areas of the globe particularly threatened by climate change such as Antarctica.

4.2. True Process: High Mitochondrial Genetic Diversity and Heteroplasmy

Although our data strongly question the validity of the extrapolation from a fragment of *COI* to the entire (nuclear) genome, the apparent division of *COI* sequences into three almost equidistant groups, namely, (h1 + h2), h3, and h4, is likely the result of a separate evolutionary history of the three in isolation from one another. The mitotype h4 is separated

by the Drake Passage and the Polar Front from its nearest relatives in the Southern Ocean, a condition that remains so until today [26]. However, divergence time estimations suggest that the haplotype h4 split off around the end of the Miocene ca. 8.5 Ma [28], thus indicating a dispersal event across the already open Drake Passage between the two continents. The differentiation of the three *Aequiyoldia* clades characterized by major haplotype groups in Antarctica likely happened in the Pliocene; however, unlike their South American relatives, all three of them occur at least in partial sympatry today. The most likely scenario for the divergent patterns found in the mitochondrial and nuclear data is that the differentiated mitochondrial lineages persisted even after the groups made secondary contact, which is a consequence of their independent, non-recombining, clonal inheritance. On the other hand, the pattern of differentiation in the nuclear genome eroded away over time as a result of the recombination of the hybridizing lineages, possibly comprising a case of speciation reversal as suggested for other Antarctic invertebrates [52]. Several examples of genetic diversification caused by temporary isolation, some of them involving cryptic speciation, were described for other Southern Ocean invertebrates such as crustaceans [53,54], polychaetes [55], and echinoderms [56], among others.

5. Conclusions

Population genetics, molecular barcoding, and molecular systematics critically depend on extrapolations made from small subsets to larger scales. Our results show that the confidence in the correctness of the extrapolation from a *COI* fragment to the entire genome as implied in molecular barcoding may not be justified when mitochondrial heteroplasmy is involved. In our study, an amplification bias with Folmer primers in the presence of two competing templates available for amplification in a heteroplasmic scenario resulted in the amplification of either of the two existing variants (Figure 5), thereby promoting erroneous inferences with high confidence. While such a volatile amplification bias with an unpredictable outcome may seem to require several conditions and hence may be assumed to be rare, it is possibly often overlooked because its evidence is often eliminated in a quality control step that ensures clean sequences preceding the analysis.

Countermeasures against this shortcoming include an active search for double peaks in chromatograms in the Sanger sequencing results, the use of haplotype-specific primers whenever there is evidence of sequence competition, and the use of alternative sequencing methods in combination with an active search for segregating sites.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes14040935/s1>, Table S1: Annotation of 18 kb- mitochondrial transcript; Table S2: *COI* p-distances; Figure S1: Shell length of *A. eightsii*.

Author Contributions: Conceptualization, M.M., L.H., D.A. and C.H.; methodology, M.M., L.H., D.A. and C.H.; formal analysis, M.M. and L.H.; investigation, M.M., D.A. and C.H.; resources, M.M., D.A. and C.H.; writing—original draft preparation, M.M. and C.H.; writing—review and editing, M.M., L.H., D.A. and C.H.; supervision, D.A. and C.H.; funding acquisition, M.M., D.A. and C.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Agency of Investigation and Innovation (ANII, Uruguay) and German Academic Exchange Service (DAAD, Germany) by PhD scholarship to M. Martínez (Grant number POS_EXT_2015_122792), and by Zempelin-Stiftung-Deutsches Stiftungszentrum (Grant number T0214/33883/2019/sm).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: *COI* amplified with Folmer and haplotype-specific primers and *histone H3* and *18S* sequences are available in GenBank (NCBI) under the accession numbers MT176797-MT176951, MT645995-MT646032, MT647927-MT647998, and MT642948-MT643019, respectively. Raw Illumina reads were deposited in the European nucleotide Archive database (EMBL-EBI) with the accessions ERR4265443 and ERR4276392–ERR4276460 under the study accession ERP122389.

Acknowledgments: The authors thank Uwe John for their advice on experimental design, Andrea Eschbach and Enrique Celemin for their technical support, and Andrea Desiderato for her valuable comments in early stages. We also thank to Instituto Antartico Chileno (INACH), Instituto Antartico Argentino (IAA), and Instituto Antartico Uruguayo (IAU) for their logistic support, to Marcelo Gonzalez (INACH), Dolores Deregibus (IAA), and Alvaro Soutullo (IAU), and Leyla Cardenas from Centro IDEAL.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hebert, P.D.N.; Cywinska, A.; Ball, S.L.; deWaard, J.R. Biological identifications through DNA Barcodes. *Proc. Biol. Sci.* **2003**, *270*, 313–321. [\[CrossRef\]](#) [\[PubMed\]](#)
2. DeSalle, R.; Goldstein, P. Review and interpretation of trends in DNA Barcoding. *Front. Ecol. Evol.* **2019**, *7*, 302. [\[CrossRef\]](#)
3. Folmer, O.; Black, M.; Hoeh, W.; Lutz, R.; Vrijenhoek, R. DNA Primers for amplification of mitochondrial Cytochrome c Oxidase Subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **1994**, *3*, 294–299. [\[PubMed\]](#)
4. Steinke, D.; Prosser, S.W.J.; Hebert, P.D.N. DNA Barcoding of marine metazoans. In *Marine Genomics*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 155–168.
5. Rollins, L.A.; Woolnough, A.P.; Fanson, B.G.; Cummins, M.L.; Crowley, T.M.; Wilton, A.N.; Sinclair, R.; Butler, A.; Sherwin, W.B. Selection on mitochondrial variants occurs between and within individuals in an expanding invasion. *Mol. Biol. Evol.* **2016**, *33*, 995–1007. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Ratnasingham, S.; Hebert, P.D. BOLD: The Barcode of Life Data System ([Http://Www. Barcodinglife. Org](http://www.barcodinglife.org)). *Mol. Ecol. Notes* **2007**, *7*, 355–364. [\[CrossRef\]](#)
7. Françoso, E.; Zuntini, A.R.; Ricardo, P.C.; Silva, J.P.N.; Brito, R.; Oldroyd, B.P.; Arias, M.C. Conserved Numts mask a highly divergent mitochondrial-COI gene in a species complex of Australian stingless bees *Tetragonula* (Hymenoptera: Apidae). *Mitochondrial DNA Part A* **2019**, *30*, 806–817. [\[CrossRef\]](#)
8. Kim, S.-J.; Lee, K.Y.; Ju, S.-J. Nuclear mitochondrial pseudogenes in *Austinochorda alayseae* hydrothermal vent crabs (Crustacea: Bythograeidae): Effects on DNA Barcoding. *Mol. Ecol. Resour.* **2013**, *13*, 781–787. [\[CrossRef\]](#)
9. Iketani, G.; Pimentel, L.; Torres, E.D.S.; do Rêgo, P.S.; Sampaio, I. Mitochondrial heteroplasmy and pseudogenes in the Freshwater prawn, *Macrobrachium amazonicum* (Heller, 1862): DNA Barcoding and phylogeographic implications. *Mitochondrial. DNA Part A* **2021**, *32*, 1–11. [\[CrossRef\]](#)
10. Mastrantonio, V.; Latrofa, M.S.; Porretta, D.; Lia, R.P.; Parisi, A.; Iatta, R.; Dantas-Torres, F.; Otranto, D.; Urbanelli, S. Paternal leakage and mtDNA heteroplasmy in *Rhipicephalus* spp. Ticks. *Sci. Rep.* **2019**, *9*, 1460. [\[CrossRef\]](#)
11. Déglétagne, C.; Abele, D.; Held, C. A Distinct mitochondrial genome with DUI-Like inheritance in the ocean quahog *Arctica islandica*. *Mol. Biol. Evol.* **2016**, *33*, 375–383. [\[CrossRef\]](#)
12. Doublet, V.; Souty-Grosset, C.; Bouchon, D.; Cordaux, R.; Marcadé, I. A thirty million year-old inherited heteroplasmy. *PLoS ONE* **2008**, *3*, e2938. [\[CrossRef\]](#)
13. Bentzen, P.; Leggett, W.C.; Brown, G.G. Length and restriction site heteroplasmy in the mitochondrial DNA of american shad (*Alosa sapidissima*). *Genetics* **1988**, *118*, 509–518. [\[CrossRef\]](#)
14. Kvist, L.; Martens, J.; Nazarenko, A.A.; Orell, M. Paternal leakage of mitochondrial DNA in the great tit (*Parus major*). *Mol. Biol. Evol.* **2003**, *20*, 243–247. [\[CrossRef\]](#)
15. Vollmer, N.L.; Viricel, A.; Wilcox, L.; Katherine Moore, M.; Rosel, P.E. The Occurrence of mtDNA heteroplasmy in multiple cetacean species. *Curr. Genet.* **2011**, *57*, 115–131. [\[CrossRef\]](#)
16. Ghiselli, F.; Maurizii, M.G.; Reunov, A.; Ariño-Bassols, H.; Cifaldi, C.; Pecci, A.; Alexandrova, Y.; Bettini, S.; Passamonti, M.; Franceschini, V.; et al. Natural heteroplasmy and mitochondrial inheritance in bivalve molluscs. *Integr. Comp. Biol.* **2019**, *59*, 1016–1032. [\[CrossRef\]](#)
17. Gusman, A.; Lecomte, S.; Stewart, D.T.; Passamonti, M.; Breton, S. Pursuing the quest for better understanding the taxonomic distribution of the system of Doubly Uniparental Inheritance of mtDNA. *PeerJ* **2016**, *4*, e2760. [\[CrossRef\]](#)
18. Zouros, E.; Oberhauser Ball, A.; Saavedra, C.; Freeman, K.R. An unusual type of mitochondrial DNA inheritance in the blue mussel *Mytilus*. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 7463–7467. [\[CrossRef\]](#)
19. Zouros, E.; Rodakis, G.C. Doubly Uniparental Inheritance of mtDNA: An unappreciated defiance of a general rule. In *Cellular and Molecular Basis of Mitochondrial Inheritance*; Springer: Berlin/Heidelberg, Germany, 2019; Volume 231, pp. 25–49. [\[CrossRef\]](#)
20. Doucet-Beaupré, H.; Breton, S.; Chapman, E.G.; Blier, P.U.; Bogan, A.E.; Stewart, D.T.; Hoeh, W.R. Mitochondrial phylogenomics of the bivalvia (Mollusca): Searching for the origin and mitogenomic correlates of Doubly Uniparental Inheritance of mtDNA. *BMC Evol. Biol.* **2010**, *10*, 50. [\[CrossRef\]](#)
21. Terranova, M.S.; Lo Brutto, S.; Arculeo, M.; Mitton, J.B. A Mitochondrial phylogeography of *Brachidontes variabilis* (Bivalvia: Mytilidae) reveals three cryptic species. *J. Zool. Syst. Evol. Res.* **2007**, *45*, 289–298. [\[CrossRef\]](#)
22. Boyle, E.E.; Etter, R.J. Heteroplasmy in a deep-Sea protobranch bivalve suggests an ancient origin of Doubly Uniparental Inheritance of mitochondria in bivalvia. *Mar. Biol.* **2012**, *160*, 413–422. [\[CrossRef\]](#)

23. Sharma, P.P.; Zardus, J.D.; Boyle, E.E.; González, V.L.; Jennings, R.M.; McIntyre, E.; Wheeler, W.C.; Etter, R.J.; Giribet, G. Into the Deep: A phylogenetic approach to the bivalve subclass protobranchia. *Mol. Phylogenetics Evol.* **2013**, *69*, 188–204. [[CrossRef](#)] [[PubMed](#)]
24. Reed, A.J.; Linse, K.; Thatje, S. Differential adaptations between cold-stenothermal environments in the bivalve *Lissarca Cf. miliaris* (Philobryidae) from the Scotia Sea islands and Antarctic Peninsula. *J. Sea Res.* **2014**, *88*, 11–20. [[CrossRef](#)]
25. González-Wevar, C.A.; Díaz, A.; Gérard, K.; Cañete, J.I.; Poulin, E. Divergence time estimations and contrasting patterns of genetic diversity between antarctic and southern South America benthic invertebrates. *Rev. Chil. Hist. Nat.* **2012**, *85*, 445–456.
26. Muñoz-Ramírez, C.P.; Barnes, D.K.A.; Cárdenas, L.; Meredith, M.P.; Morley, S.A.; Roman-Gonzalez, A.; Sands, C.J.; Scourse, J.; Brante, A. Gene flow in the antarctic bivalve *Aequiyoldia eightsi* (Jay, 1839) suggests a role for the Antarctic Peninsula coastal current in larval dispersal. *R. Soc. Open Sci.* **2020**, *7*, 200603. [[CrossRef](#)]
27. Martínez, M.; González-Aravena, M.; Held, C.; Abele, D. A molecular perspective on the invasibility of the Southern Ocean benthos: The impact of hypoxia and temperature on gene expression in South American and Antarctic *Aequiyoldia* bivalves. *Front. Physiol.* **2023**, *14*, 269.
28. González-Wevar, C.A.; Gérard, K.; Rosenfeld, S.; Saucède, T.; Naretto, J.; Díaz, A.; Morley, S.A.; Brickle, P.; Poulin, E. Cryptic speciation in Southern Ocean *Aequiyoldia eightsi* (Jay, 1839): Mio-Pliocene trans-Drake Passage separation and diversification. *Prog. Oceanogr.* **2019**, *174*, 44–54. [[CrossRef](#)]
29. Low, T.Y.; Mohtar, M.A.; Ang, M.Y.; Jamal, R. Connecting proteomics to Next-Generation Sequencing: Proteogenomics and its current applications in biology. *Proteomics* **2019**, *19*, e1800235. [[CrossRef](#)]
30. De Wit, P.; Pespeni, M.H.; Palumbi, S.R. SNP genotyping and population genomics from expressed sequences—Current advances and future possibilities. *Mol. Ecol.* **2015**, *24*, 2310–2323. [[CrossRef](#)]
31. García, G.; Ríos, N.; Gutiérrez, V.; Serra, S.; Loureiro, M. Transcriptome-based SNP discovery and validation in the hybrid zone of the neotropical annual fish Genus *Austrolebias*. *Genes* **2019**, *10*, 789. [[CrossRef](#)]
32. Yan, J.; Song, Z.; Xu, Q.; Kang, L.; Zhu, C.; Xing, S.; Liu, W.; Greimler, J.; Züst, T.; Li, J.; et al. Population transcriptomic characterization of the genetic and expression variation of a candidate progenitor of *Miscanthus energy* Crops. *Mol. Ecol.* **2017**, *26*, 5911–5922. [[CrossRef](#)]
33. Rogier, O.; Chateigner, A.; Amanzougarene, S.; Lesage-Descauses, M.-C.; Balzergue, S.; Brunaud, V.; Caius, J.; Soubigou-Taconnat, L.; Jorge, V.; Segura, V. Accuracy of RNAseq based SNP discovery and genotyping in *Populus nigra*. *BMC Genom.* **2018**, *19*, 909. [[CrossRef](#)]
34. Bushnell, B. BBMap. BBMap Short Read Aligner, and Other Bioinformatic Tools. 2014. Available online: <https://sourceforge.net/projects/bbmap/> (accessed on 24 July 2021).
35. Kopylova, E.; Noé, L.; Touzet, H. SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **2012**, *28*, 3211–3217. [[CrossRef](#)]
36. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [[CrossRef](#)]
37. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359.
38. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
39. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [[CrossRef](#)]
40. O’Leary, S.J.; Puritz, J.B.; Willis, S.C.; Hollenbeck, C.M.; Portnoy, D.S. These aren’t the loci you’re looking for: Principles of effective SNP filtering for molecular ecologists. *Mol. Ecol.* **2018**, *27*, 3193–3206. [[CrossRef](#)]
41. Bernt, M.; Donath, A.; Jühling, F.; Externbrink, F.; Florentz, C.; Fritzsch, G.; Pütz, J.; Middendorf, M.; Stadler, P.F. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenetics Evol.* **2013**, *69*, 313–319. [[CrossRef](#)]
42. Verity, R.; Nichols, R.A. What is genetic differentiation, and how should we measure It—GST, D, neither or both? *Mol. Ecol.* **2014**, *23*, 4216–4225. [[CrossRef](#)]
43. Excoffier, L.; Smouse, P.E.; Quattro, J.M. Analysis of Molecular Variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **1992**, *131*, 479–491. [[CrossRef](#)]
44. RCore, T. R: A Language and Environment for Statistical Computing; R foundation for statistical computing: Vienna, Austria, 2016.
45. Knaus, B.J.; Grünwald, N.J. Vcfr: A package to manipulate and visualize variant call format data in R. *Mol. Ecol. Resour.* **2017**, *17*, 44–53. [[CrossRef](#)] [[PubMed](#)]
46. Jombart, T.; Ahmed, I. Adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* **2011**, *27*, 3070–3071. [[CrossRef](#)] [[PubMed](#)]
47. Kamvar, Z.N.; Tabima, J.F.; Grünwald, N.J. Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2014**, *2*, e281. [[CrossRef](#)] [[PubMed](#)]
48. Dray, S.; Dufour, A.-B. The Ade4 package: Implementing the duality diagram for ecologists. *J. Stat. Softw.* **2007**, *22*, 1–20. [[CrossRef](#)]
49. Paradis, E. Pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics* **2010**, *26*, 419–420. [[CrossRef](#)]

50. Milani, L.; Ghiselli, F. Faraway, so close. The comparative method and the potential of non-model animals in mitochondrial research. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2020**, *375*, 20190186. [[CrossRef](#)]
51. David, K.T.; Wilson, A.E.; Halanych, K.M. Sequencing disparity in the genomic era. *Mol. Biol. Evol.* **2019**, *36*, 1624–1627. [[CrossRef](#)]
52. Dietz, L.; Arango, C.P.; Dömel, J.S.; Halanych, K.M.; Harder, A.M.; Held, C.; Mahon, A.R.; Mayer, C.; Melzer, R.R.; Rouse, G.W.; et al. Regional differentiation and extensive hybridization between mitochondrial clades of the Southern Ocean giant sea spider *Colossendeis megalonyx*. *R. Soc. Open Sci.* **2015**, *2*, 140424. [[CrossRef](#)]
53. Baird, H.P.; Miller, K.J.; Stark, J.S. Evidence of hidden biodiversity, ongoing speciation and diverse patterns of genetic structure in giant antarctic amphipods. *Mol. Ecol.* **2011**, *20*, 3439–3454. [[CrossRef](#)]
54. Held, C. Molecular evidence for cryptic speciation within the widespread antarctic crustacean *Ceratoserolis trilobitoides* (Crustacea, Isopoda). In *Antarctic Biology in A Global Context*; Gregor Rare Books: Langley, WA, USA, 2003; pp. 135–139. ISBN 90-5782-079-X.
55. Neal, L.; Wiklund, H.; Muir, A.I.; Linse, K.; Glover, A.G. The identity of juvenile Polynoidae (Annelida) in the Southern Ocean revealed by DNA taxonomy, with notes on the status of *Herdmanella gracilis ehlers* Ssensu augener. *Mem. Mus. Vic.* **2014**, *71*, 203–216.
56. Hemery, L.; Eléaume, M.; Roussel, V.; Améziane, N.; Gallut, C.; Steinke, D.; Cruaud, C.; Couloux, A.; Wilson, N. Comprehensive sampling reveals circumpolarity and sympatry in seven mitochondrial lineages of the Southern Ocean crinoid species *Promachocrinus kerguelensis* (Echinodermata). *Mol. Ecol.* **2012**, *21*, 2502–2518. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.