Citation:

William P. Goodall-Copestake; One tunic but more than one barcode: evolutionary insights from dynamic mitochondrial DNA in *Salpa thompsoni* (Tunicata: Salpida). *Biol J Linn Soc* 2017; 120 (3): 637-648. doi: 10.1111/bij.12915

The published version can be found at:

https://academic.oup.com/biolinnean/article/120/3/637/3056001/One-tunic-but-more-than-one-barcode-evolutionary

TITLE

One tunic but more than one barcode: evolutionary insights from dynamic mitochondrial DNA in the plankton *Salpa thompsoni* (Salpida)

AUTHOR

William Paul Goodall-Copestake

INSTITUTION

British Antarctic Survey (Natural Environment Research Council), High Cross, Madingley Road, Cambridge, CB3 0ET, UK.

Tel.: +44 (0)1223 221653

Fax: +44 (0)1223 362616

Email: wgco@bas.ac.uk

RUNNING TITLE

Dynamic barcodes in salps

ABSTRACT

The DNA barcode within the mitochondrial *cox1* gene is typically used to assess the identity and diversity of animals under the assumption that individuals contain a single form of this genetic marker. This study reports on a novel exception from the pelagic tunicate *Salpa thompsoni* Foxton. Oozooids caught off South Georgia and the Antarctic Peninsula generated barcodes consisting of a single prominent DNA sequence with some additional, subtler signals of intra individual variation. Further investigation revealed that this was due to duplicated and/or minicircular DNAs. These could not simply be explained as artefacts or nuclear copies of mitochondrial DNA, but provided evidence for heteroplasmy arising from a dynamic mitochondrial genome. Genetic variation of this sort may enable *S. thompsoni* to ecologically benefit from asexually-driven population blooms without incurring the genetic cost of excessive mutational load. Analysis of the prominent barcode sequence data yielded low haplotype ($h < 0.61$) and nucleotide ($\pi < 0.0014$) diversities, and no evidence for genetic structure between sampling locations as assessed using AMOVA. These results are consistent with the impact of both population blooms and the mixing effect of Southern Ocean currents on *S. thompsoni* genetic diversity.

ADDITIONAL KEYWORDS

Antarctica - COI - *cox1* - heteroplasmy - minicircles - recombination - salps - sublimons - tunicate

INTRODUCTION

The cytochrome c oxidase subunit I (*cox1*, COI) DNA barcode, championed by Hebert *et al.* (2003), has gained a prominent role in biodiversity research as often the first and sometimes the only genetic marker used to assess the identity and diversity of animal species. The default expectation for this molecular marker is that it is a single mitochondrial locus from which a single DNA sequence will be generated. Many results from *cox1* sequencing surveys are consistent with this simple scenario (e.g. Hebert *et al.*, 2003); however, it has long been acknowledged that the biology of mitochondrial DNA is more complicated (Bensasson *et al.*, 2001; Birky, 2001), and as such, animals can and do contain more than one DNA barcode.

Intra individual barcode variation is usually attributed to duplicates of mitochondrial DNA within the nuclear genome (numts; reviewed in Bensasson *et al.*, 2001). Variable *cox1* containing numts have now been characterised from the sequenced nuclear genomes of several species (Hazkani-Covo, Zeller & Martin, 2010), among them the honeybee *Apis mellifera* L., which harbours more numts from *cox1* than any other mitochondrial region (Behura, 2007). DNA duplications also occur within animal mitochondrial genomes, although reports of these are rare compared to numts. *Cox1* duplications are known from squid (e.g. *Watasenia scintillans* Berry; Yokobori *et al.*, 2004) and the linear mitochondrial genomes of nontrachyline hydrozoans (Kayal *et al.*, 2012).

Heteroplasmy - the presence of more than one type of mitochondrial DNA in an individual (reviewed in Kmiec, Woloszynska & Janska, 2006) - represents another, rarely used, explanation for intra individual barcode variation. Heteroplasmy can arise when multiple mitochondrial haplotypes are inherited (Kmeic *et al.,* 2006). A classic example of this is doubly uniparental inheritance in mussels, which was shown to generate intra individual *cox1* variability in *Mytilus* L. species by Venetis *et al.* (2006). Heteroplasmy can also arise *de novo* (typically as somatic variation) through recombination and small-scale mutations derived from replication and repair errors, as well as from the activity of reactive metabolites, in both large genomic and smaller fragments of mitochondrial DNA known as minicircles and sublimons (Kajander *et al.*, 2000; Kmeic *et al.,* 2006; Mao *et al.,* 2014). This type of heteroplasmy is not usually acknowledged as a source of intra individual *cox1* variation, although Herd, Barker & Shao (2012) proposed that recombination generated heteroplasmy among *cox1* containing mini chromosomes within the multi-partite louse (*Pediculus* L.) mitochondrial genome.

Both numts and heteroplasmy are considered to be more widespread in animals than previously assumed (Kmeic *et al.*, 2006; Hazkani-Covo *et al.*, 2010). This may seem at odds with the results from *cox1* sequencing surveys that usually present no evidence for numts or heteroplasmy. However, the majority of *cox1* sequencing surveys published to date involved

directly Sanger sequencing (Sanger, Nicklen & Coulson, 1977) PCR amplicons, a procedure that will not reveal the presence of multiple amplicon variants unless these are similar in abundance, e.g. when one variant is no more than five times more abundant than another (Wong & Boles, 2005). Mitochondrial *cox1* is usually expected to far outnumber *cox1*-bearing numts, and most cases of heteroplasmy are believed to involve just a single prominent form of mitochondrial DNA with the other variants present at lower levels (Frey & Frey, 2004; Kmeic *et al.*, 2006). Consequently, if an individual does contain multiple *cox1* barcodes, these may not be discovered through direct sequencing due to the limited resolution of this method - assuming all such variants are amplified with a similar efficacy, which is not always the case (c.f. Bensasson *et al.*, 2001).

Salps are pelagic filter feeders with alternating sexual and asexual generations (blastozooids and oozooids), the latter of which facilitate the development of salp blooms under favourable conditions (Foxton, 1961; Daponte, Capitanio & Esnal, 2001; Loeb & Santora, 2012). Salp blooms can have a large impact on the structure of marine communities, and yet despite their ecological importance, salps have only recently been subject to *cox1* barcode sequencing in a study involving two species from the genus *Thalia* Blumenbach (Goodall-Copestake, 2014). For the present study, fragments of nuclear ribosomal small subunit (18S) DNA and *cox1* were directly Sanger sequenced to investigate the most abundant salp in the Southern Ocean, *Salpa thompsoni* Foxton (Foxton, 1961). Base calling within the *cox1* sequences obtained was unambiguous, although some sites also had subtle alternative base calls. This was consistent with the amplification of a single predominant form of DNA barcode accompanied by lower abundance barcode variation and/or methodological artefacts. Further investigation using cloned PCR amplicons and inverse PCR uncovered numerous barcode variants within rearranged fragments of mitochondrial DNA. The origin of this intra individual variation and its evolutionary significance is discussed. Inter individual variation was also assessed by analyzing sequence data corresponding to the predominantly amplified form of DNA barcode. The resulting diversity estimates for *S. thompsoni* are considered within the context of standardized estimates from other animal species.

MATERIAL AND METHODS

SPECIMEN SELECTION

Plankton samples from the British Antarctic Survey -80°C archives of oceanographic cruise JR26 (November-December 1997) were defrosted and sorted to select for *S. thompsoni* oozooids following the species description in Foxton (1961). Salp oozooids are considerably less abundant than blastozooids within plankton samples (Foxton, 1961; Daponte *et al.*, 2001; Loeb & Santora, 2012). However, despite the challenge of obtaining oozooids, individuals from this life stage were

specifically targeted to provide a direct measure of the products of sexual reproduction, rather than blastozooids which may be identical due to their asexual origin. Two of the pelagic sampling hauls from cruise JR26 contained a sufficient number of oozooids for population-level diversity analysis (c.f. Goodall-Copestake, Tarling & Murphy, 2012). Twenty-four oozooids were sub-sampled from each of these collections: a single net haul taken at approximately 55S 41W on the shelf break west of South Georgia (samples e146 1-24), and a single haul at 62S 56W on the northern shelf off the Antarctic Peninsula (samples e187 1-24).

MOLECULAR DATA GENERATION

Muscle bands were dissected from the 48 *S. thompsoni* oozooids and DNA was extracted using a DNeasy Tissue Kit (Qiagen). Fragments of 18S and the *cox1* barcode were PCR amplified using a Velocity polymerase mix (Bioline) in 15 μL reactions containing c.10 ng of template and a final concentration of 0.2 μM of each primer: 18S F and 18S R for 18S, LCO variant and HCO variant for *cox1* (Table 1). The thermocycling profile was 60 s at 98°C then 30 cycles of 10 s at 98°C, 10 s at 62°C (for 18S) or 56°C (for *cox1*), and 90 s at 72°C (ramping to 72°C at 0.3°C/s). The primers 18S F and 18S R were designed for this study from an alignment of the sequence data in Govindarajan, Bucklin & Madin (2011) to amplify a 563 bp species-specific sequence. The degenerate primers LCO variant and HCO variant (Goodall-Copestake, 2014) were used to amplify *cox1* because the popular invertebrate primers LCO and HCO (Folmer *et al.*, 1994) failed to work. PCR amplicons were verified through visualization under UV light on GelRed (Biotium) stained 1.5% agarose gels and subsequently cleaned for unincorporated reagents using Illustra ExoProStar (GE Healthcare). All amplicons were bi-directionally Sanger sequenced at commercial facilities (Source BioScience, Eurofins) using the original amplification primers for 18S and the sequencing oligonucleotides TAILED and PRIMER for *cox1* (Goodall-Copestake, 2014; Table 1).

The directly sequenced *cox1* amplicons generated electropherograms that contained sites where a second weak peak occurred under the main prominent base call. For this reason, and because degenerate *cox1* primers were used, six individuals (e146-2, e146-11, e146-19, e187-3, e187-9, e187-19) were selected for further analysis. *Cox1* amplicons from these individuals were modified by adding 3'A overhangs using a PCR tailing mix (Bioline) and subsequently cloned using a TOPO TA Cloning Kit (Life Technologies). Eight clones from each reaction were then bi-directionally sequenced using the M13 primers supplied with the cloning kit.

Some of the intra individual *cox1* amplicons showed more differences than expected, and consequently, *cox1*-bearing DNAs from all six individuals listed above were further investigated using a barcode-primed inverse PCR approach. This method was chosen to exploit data-in-hand on salp mitochondrial DNA and thereby avoid complications with the use of multiple combinations of

degenerate primers. It was based on the assumption that salp mitochondrial genomes have a typical circular form of c.15 Kbp like other tunicates (Gissi, Iannelli & Pesole, 2008), and so inverse PCR could be used to obtain sequence data from regions flanking the DNA barcode. A pair of inverse PCR primers were designed with a 3 bp 5' overlap to an invariant region of *S. thompsoni cox1* DNA (Inverse F and Inverse R, Table 1, Figure 1) and used at a final concentration of 0.2 μM in 25 μL long-range PCR reactions containing c.15 ng of template and a Ranger DNA Polymerase mix (Bioline). The thermocycling profile was 60 s at 95°C, then 30 cycles of 10 s at 98°C and 12 min at 62°C. The resulting PCR amplicons, which were dominated by 2.0-2.5 Kbp fragments, were isolated using the standard cloning procedure described above. Eight cloned fragments that appeared to have different insert sizes were selected from each cloning reaction and fully bi-directionally sequenced using M13 primers and four internal sequencing primers (listed as 'IntSeq' primers in Table 1) to generate the 'cloned inverse PCR amplicon sequences'.

DNA SEQUENCE CHARACTERIZATION

Electropherograms were edited and assembled into contigs using the software Geneious v6 (Biomatters: http://www.geneious.com). The 18S sequences were used for a molecular corroboration of species identity by direct comparison with the *S. thompsoni* sequences from Govindarajan *et al.* (2011). The *cox1* containing sequences were characterised by conducting BLAST searches of NCBI archives (http://blast.ncbi.nlm.nih.gov; Johnson *et al.*, 2008) and by translating the open reading frames associated with BLAST gene hits into amino-acids using the ascidian mitochondrial code. The software Arwen v1.2 (Laslett & Canbäck, 2008) was also used to search for mitochondrial tRNAs.

ANALYSING INTRA INDIVIDUAL DNA VARIATION

All of the *cox1* degenerate PCR amplicon sequences (obtained by direct sequencing and from cloned amplicons) were manually aligned and compared to assess for intra individual barcode variation. Unique sequences from this *cox1* amplicon alignment were subsequently aligned with the cloned inverse PCR amplicon sequences. This was straightforward and unambiguous except over a few sites either side of a putative inversion and deletion within a short stretch of NADH dehydrogenase (*nad4*) coding DNA. Manual inspection of the alignment revealed adjacent regions of conflicting nucleotide site patterns that clearly supported different sequence groupings. To accommodate and explore these conflicting site patterns, variation among the aligned sequences was graphically visualized by generating a splits network. For this, sequences containing long gaps, gap containing sites, and the ambiguously aligned region containing an inversion and deletion were removed from the alignment to generate a matrix containing 29 sequences that was 1884 bp long.

Uncorrected pairwise distances calculated from this matrix were analysed using the Neighbour-Net method (Bryant & Moulton, 2004) with the software Splitstree v4.14.2 (Huson & Bryant, 2006). Sliding-window phylogenetic analysis was also used to assess the most obvious site pattern conflicts within the DNA barcoding region. Four inverse PCR product sequences derived from the same individual (e187-3) were used for this to ensure that the results obtained could be attributed to intra rather than inter individual variation. The sliding window phylogenetic analysis was conducted on a 536 bp partition of *cox1* using the Geneious Dualbrothers recombination detection plugin v1.1.3 (Minin *et al.*, 2005), under default settings for a comparison of all possible topologies.

ASSESSING PCR ARTEFACTS AS SOURCES OF VARIATION

DNA extension during PCR amplification can be interrupted when template re-annealing occurs between highly similar stretches of DNA (Saiki *et al.*, 1988; Meyerhans, Vatanian & Wain-Hobson, 1990). This 'discontinuous DNA extension' can generate sequences containing artificial gaps and also mosaics when variable templates are present. Two series of tests were conducted to investigate the possibility that such artefacts might have generated the intra individual DNA variation found in this study.

The first test was similar to one recently described by Mao *et al.* (2014) and involved investigating whether a single cloned mitochondrial DNA template could generate shorter amplicon artefacts during PCR (see online Supporting Information for a graphical representation). This involved repeating the degenerate and inverse PCR reactions using the original 735 bp (degenerate) and 2,501 bp (inverse) cloned sequences as templates, respectively. A semi-quantitative, high cycle number PCR method was used to increase the possibility of detecting artefacts and to place the original 30 cycle PCRs into context: 75 μL reactions were prepared and split into five aliquots of 15 μL, one of these aliquots was removed from the thermocycler after 25, 30, 35, 40 and 45 cycles. The resulting amplicons were compared under UV on GelRed stained agarose gels.

The second PCR test aimed to assess if discontinuous DNA extension might account for mosaic-like amplicons (see Supporting Information). Genomic DNA from individual e187-3 was used as a template for this because it showed strong evidence of mosaic *cox1* sequences. Modified versions of the inverse PCR primers were designed that contained 18 bp 5' extensions matching TAILED, PRIMER and a novel 18-MER sequence (Table 1). Using different combinations of these primers and the same conditions as in the original inverse PCR, three sets of amplicons were generated in which the first and last 18 bp were: TAILED...18-MER (A amplicons), 18-MER...PRIMER (B amplicons), and TAILED...PRIMER (C amplicons). Unincorporated primers were removed using Illustra ExoProStar and the amplicons were then diluted into 10,000 volumes of water. Diluted amplicons A, B, C and a mixture of A + B were then used as templates for a

second round of PCR, this time under the conditions described above for the semi-quantitative PCR. Three different combinations of primers were used for these PCR reactions: TAILED + 18-MER, 18-MER + PRIMER, and TAILED + PRIMER. If mosaics were artificially generated, then primer combination TAILED + PRIMER would be expected to exponentially amplify the A + B template. In the absence of mosaic generation, only primer combinations TAILED + 18-MER and 18-MER + PRIMER would be expected to exponentially amplify the A + B template. The other templates (A, B, C) were used for positive and negative control reactions.

INTER-INDIVIDUAL ANALYSIS OF DNA BARCODES

The strength of the peaks in electropherograms obtained by directly sequencing *cox1* degenerate PCR amplicons, and the low frequency with which cloned variants of these amplicons were recovered at the intra-individual, suggested that the degenerate *cox1* primers predominantly amplified a single form of DNA barcode from each oozooid used in this study. These barcodes were considered as 'representative' of individuals for the purposes of population-level analysis. For this, the *cox1* degenerate PCR amplicon alignment was split into two groups representing each sampling event (geographic origin). The software Arlequin v.3.1 (Excoffier, Laval & Schneider, 2005) was then used to calculate descriptive statistics for each group and to conduct non-hierarchical analysis of molecular variance (AMOVA) based on distances calculated from the raw number of sequence differences. Significance values were generated using 10,000 non-parametric permutations of the data set. Descriptive statistics were also calculated from a 456 bp partition of the *cox1* alignment to provide a direct comparison with the standardised *cox1* population-level diversity estimates in Goodall-Copestake *et al.* (2012).

RESULTS

DNA SEQUENCES

Forty-eight *S. thompsoni* 18S sequences were generated, each of which was 516 bp long (563 bp with primers) and identical. These sequences corroborated the morphological species designation as they showed 100% similarity to the corresponding 516 bp region of 18S in the *S. thompsoni* GenBank accessions HQ015406 and FM244867. The 48 electropherograms obtained by directly sequencing *cox1* degenerate PCR amplicons contained no forward and reverse read conflicts, and while they were generally clean, close inspection revealed evidence of either template heterogeneity or sequence reaction error (a second weaker peak under the main prominent base call at some sites). These c*ox1* sequences were 646 bp long (735 bp with primers) and contained no frame shifts, stop codons or differences after amino acid translation. Ninety-six degenerate PCR and inverse PCR

cloned amplicon sequences were generated. These sequences were all unambiguous and ranged from 278-2452 bp (327-2501 bp with primers). All of the sequence data, except that from cloned degenerate *cox1* PCR amplicons that was identical to direct amplicon sequences from the same individual, were submitted to the DNA Databank of Japan (DDBJ) and assigned the accession numbers LC011709-LC011855.

INTRA INDIVIDUAL DNA VARIATION

Intra individual variation was not detected among the eight cloned degenerate PCR amplicons sequenced from individuals e146-11, e146-19 and e187-3. However, individual e146-2 had a single clone that differed at 22 sites from the other seven clones obtained from this individual. For individual e187-19, a single clone was found that differed by a single site from the other e187-19 clones. Individual e187-9 produced one clone with a 324 bp deletion that otherwise showed no sequence differences from the other e187-9 clones. These results provided evidence for multiple forms of *cox1* DNA in *S. thompsoni* that occurred at a low, but non-negligible level within the degenerate PCR amplicon pool.

Inverse PCR did not recover any observable high molecular weight amplicons consistent with a typical c.15 Kbp tunicate mitochondrial genome despite the 12-minute extension time. Instead, a range of smaller 0.3-2.5 Kbp amplicons were obtained. Each oozooid generated a unique profile of bands within this size range, with those 2.0-2.5 Kbp dominating the amplicon pools. These profiles were reproducible as they were consistently generated in replicate PCRs. The smallest, 327 bp amplicon, consisted solely of *cox1* DNA. After taking into account the 3 bp 5' overlap between the inverse PCR primers, it was clear that this 327 bp fragment exactly matched the 324 bp deletion in the cloned degenerate PCR amplicon from individual e187-9. The boundaries of this deleted region contained 15 bp direct sequence repeats (Figure 1). Inverse PCR amplicons longer than 327 bp were comprised of varying amounts of *cox1* with other mitochondrial functional DNAs including the NADH dehydrogenase subunit 3 (*nad3)* and subunit 4 (*nad4)* genes, and Met (CAT) and Ser (ACT) tRNAs (Figure 2). A comparison of the shorter amplicons with longer ones revealed 5-15 bp direct sequence repeats that bordered all of the deletions in the smaller molecules, except in one deletion that neighboured an inversion within a short stretch of ambiguously aligned *nad4* DNA. All of the inverse PCR amplicons could occur *in vivo* as single circular DNAs and/or as tandem repeats within a larger circular or linear molecule (Figure 2).

Sequence differences between the cloned inverse PCR amplicons were concentrated within the truncated 3' end of *cox1* and adjacent *nad3* gene, as well as within the ambiguously aligned area of inverted DNA within the *nad4* gene. Differences within the remainder of the alignment amounted to little more than a scattering of variable sites. The splits network separated most of the

cloned inverse PCR amplicons into two clusters that reflected the main *cox1-nad3* sequence differences (Figure 3). Uncorrected pairwise differences between these two clusters, calculated using 512 bp of *cox1* extending from the 3' end of the LCO variant primer, were as high as 8% and the corresponding translated (amino acid) differences up to 20%. These results suggested that at least part of the mitochondrial genome occurs as multiple, heritable, copies in *S. thompsoni* and that these have diverged under different protein coding constraints. It was notable that the *cox1* degenerate PCR primers predominantly amplified just a single variant of these mitochondrial DNA duplicates (indicated by asterisks in Figure 3). The most obvious site-pattern conflicts were associated with cloned inverse PCR amplicon sequences that occurred on splits between the two main clusters within the network (indicated by grey squares in Figure 3). Among these, sequences from individual e187-3 showed some of the strongest such conflicts. Sliding window phylogenetic analysis of this e187-3 sequence data produced alternative, well-supported topologies either side of a break point within the *cox1* barcode (Figure 4). This result provided convincing evidence for *in vivo* or *in vitro* recombination between *S. thompsoni cox1* DNAs. In addition to the most obvious sequence differences and site pattern conflicts represented in the network (3' end of *cox1 - nad3* region), there were other subtler variations within the inverse PCR amplicon alignment that were represented by splits (Figure 3). Some of this variation may also be derived from DNA duplications and recombination, or it may have arisen from other biological sources, and possibly also methodological artefacts.

POTENTIAL FOR PCR ARTEFACTS

Reactions using the semi-quantitative degenerate PCR method with 735 bp cloned sequences as templates did not yield any perceivable amplicons <735 bp after 45 cycles of PCR. However, reactions using the semi-quantitative inverse PCR method with 2,501 bp cloned sequence templates generated several discrete amplicons <2,501 bp after 45 cycles. Genomic DNA included as an inverse PCR positive control yielded 0.3-2.5 Kbp amplicon profiles after just 25 PCR cycles. Reactions using the semi-quantitative inverse PCR method with the mixed A + B template and primer combination TAILED + PRIMER only generated perceivable amplicons after 45 PCR cycles; these were in the 2.0-2.5 Kbp size range. By contrast, the same template amplified with either of the primer combinations TAILED + 18-MER or 18-MER + PRIMER (i.e. positive control conditions) generated a 0.3-2.5 Kbp product profile after 25 cycles. Taken together, these PCR test results demonstrated that artefacts can arise, but crucially, these only comprised a minor component after 30 PCR cycles, which was the amplification stage assessed for sequence variation during the original PCRs.

INTER-INDIVIDUAL BARCODE VARIATION

The 48 electropherograms obtained by directly sequencing *cox1* degenerate PCR amplicons comprised 12 unique haplotypes, two of which were shared between the two population-level samples. The frequencies of these shared haplotypes were fairly similar in both samples: 0.750 and 0.625 (shared haplotype 1), and 0.083 and 0.125 (shared haplotype 2) in e146 (South Georgia) and e187 (Antarctic Peninsula), respectively. Haplotype diversity and nucleotide diversity within both localities was similar and low (Table 2). Non-hierarchical AMOVA analyses revealed more of this diversity was within rather than between locations as ϕ_{ST} was negative and not significant ($\phi_{ST} = -0.00355$, $P = 0.66855$). The AMOVA analysis thus provided no evidence for mitochondrial DNA structure between the *S. thompsoni* oozooids sampled from South Georgia and the Antarctic Peninsula.

DISCUSSION

The sequence data presented in this study unambiguously demonstrated that individual *S. thompsoni* oozooids contain multiple variants of the *cox1* barcode and neighbouring regions of mitochondrial DNA (Figures 1-3). This intra individual DNA variation could have a genuine *in vivo* origin and/or it might reflect methodological *in vitro* artefacts. Artificial sources of *cox1* variation introduced by cloning and Sanger sequencing have been assessed previously and would only be expected to affect a few sites at most (Frey & Frey, 2004; Goodall-Copestake *et al.*, 2010). Therefore, these sources of error could not account for the vast majority of intra individual mitochondrial DNA differences found. Discontinuous DNA extension following interrupted template re-annealing during PCR (Saiki *et al.*, 1988; Meyerhans *et al.*, 1990) could have had a larger impact through the generation of amplicon length variants and chimeric amplicons containing site pattern conflicts. Results from the PCR tests demonstrated that these artefacts can indeed occur, but importantly, their accumulation during PCR was low relative to genuine target amplicons even after 45 cycles. Thus, while these artefacts cannot be ruled out as a source for amplicon variation, it seems reasonable to conclude that most of the intra individual mitochondrial DNA differences found in this study appear to have had a genuine biological origin.

Numts within the nuclear genome seem to provide a logical *in vivo* source for intra individual mitochondrial DNA variation in *S. thompsoni*. They can show varying levels of divergence, are recombining, and evolve in the absence of protein coding constraints (Bensasson *et al.*, 2001; Hazkani-Covo *et al.*, 2010). However, in order to explain the numerous variable-length inverse PCR amplicons recovered (Figure 2), it is necessary to assume PCR priming off multiple repeats of numt deletion variants within the *S. thompsoni* nuclear genome. Notwithstanding the difficulties associated with correctly assembling repetitive DNA from short sequence reads (Nagarajan & Pop,

2013), examples of such complex templates have yet to be described from genome-wide characterizations of numts (Hazkani-Covo *et al.*, 2010). Therefore, numts do not appear to provide an adequate explanation for all of the intra individual mitochondrial DNA variation found here.

An alternative *in vivo* source for this intra individual DNA variation is one or more duplications within the mitochondrial genome - involving relaxed coding constraints - and heteroplasmy generated from this *de novo* in the form of minicircles. Duplications within the mitochondrial genome are certainly plausible given examples from other animals, e.g. squid (Yokobori *et al.*, 2004). Some of the inverse PCR amplicons may have directly primed off such duplicated DNA. The remaining inverse PCR amplicons, perhaps even all of them, can be explained as having primed off DNA minicircles. Mitochondrial DNA minicircles are known to contain short direct repeats at inferred break points, signatures of recombination, point mutations, and they vary in abundance with tissue type and individual (Kajander *et al.*, 2000; Kmeic *et al.,* 2006). Minicircles are usually thought to occur at sub-stoichiometric levels, however, in oxidatively stressed tissues like muscles they may comprise a non-negligible fraction of total DNA that will be preferentially amplified over larger mitochondrial molecules during long-range PCR (Kajander *et al.*, 2000). The similarity of the inverse (long-range) PCR amplicons recovered from *S. thompsoni* in this study to previous descriptions of minicircles is striking, and the potential for their ready amplification at different stoichiometries from muscle DNA extractions is convincing.

PCR artifacts, numts, and duplication-minicircle derived heteroplasmy may all have contributed to the intra individual DNA variation described herein, although it is the latter of these possibilities that represents the most parsimonious explanation. This interpretation of a ‘dynamic mitochondrial genome’ is of potential functional significance with respect to both the life history of salps and the evolution of tunicate mitochondrial DNA. Key features of salp life history include the capacity to rapidly form blooms, which has a considerable population subsampling effect, and the ability to asexually generate hundreds of blastozooid offspring (Foxton, 1966; Daponte *et al.*, 2001). The combined impact of both of these processes would be expected to exacerbate mutational load within *S. thompsoni* mitochondrial lineages, potentially resulting in a mutational meltdown (Gabriel, Lynch & Bürger, 1993). A dynamic mitochondrial genome would provide the mechanisms by which this mutational load could be purged and possibly even exploited to facilitate adaptation. Indeed, this may help to explain how salps have evolved to benefit from a life history that is characterized by opportunistic, asexually driven population blooms without incurring the ultimate cost of a mutational meltdown. Over longer evolutionary timescales, the processes of DNA duplication and minicircle formation have also been considered to be important in shaping animal mitochondrial genomes (Gissi *et al.,* 2008; Mao *et al.*, 2014). Compared to many other animal subphyla, tunicates are reported to show high levels of structural variability and divergence in their

mitochondrial DNA (Gissi *et al.*, 2008; Singh *et al.*, 2009). The inference of DNA duplication and minicircle formation in this study on the pelagic tunicate *S. thompsoni*, may therefore, also help to explain these deeper evolutionary differences between tunicate mitochondrial genomes.

The discovery of multiple intra individual *cox1* barcodes might seem to negate the use of barcode DNA as a molecular marker for *S. thompsoni*. However, even though multiple barcode amplicons were discovered, only a single prominent barcode profile was obtained from each individual oozooid using standard sequencing methods. This allowed for the assembly of a 'representative' dataset for biodiversity analyses. Such an approach has been taken previously (Frey & Frey, 2004) and is accepted as standard practice when sequencing multi-copy ribosomal DNA that may not have been completely homogenized through concerted evolution (Arnheim *et al.*, 1980). Population-level analysis of the barcode data generated lower diversity estimates (Table 2) than the median diversity values given in a previous standardised comparison of animal barcodes (h = 0.70130; π = 0.00356; Goodall-Copestake *et al.*, 2012). Marine zooplankton can attain large population sizes and typically show higher levels of genetic diversity. The only published estimates of barcode diversity for salp oozooids, obtained from a sample of *Thalia longicauda* Quoy & Gainmard using the same degenerate PCR primers as used in this study (h = 0.9739, π = 0.00712; Goodall-Copestake, 2014) were also higher than the values for *S. thompsoni*. This raises the question as to why the *S. thompsoni* samples examined here had comparatively low haplotype and nucleotide diversities. The oozooids of *S. thompsoni* contain more developing blastozooids than the oozooids of *T. longicauda* (pers. obs.). All other things being equal, this would imply that *cox1* diversity should be lower in *S. thompsoni*. Differences between *S. thompsoni* and *T. longicauda* in terms of their population demographic histories is another potential explanation. Population blooms in *S. thompsoni* can be massive, particularly during warm 'salp years' (Daponte *et al.*, 2001). A history of such blooms in *S. thompsoni* would be expected to generate lower diversity estimates in this species. Large *S. thompsoni* aggregations have been described from waters off the Antarctic Peninsula (Loeb & Santora, 2012), and salps from these could readily have been dispersed to South Georgia given the overriding pattern of ocean currents in the region (Thorpe *et al.*, 2004). Such a scenario would not only help to explain the low barcode diversity at both locations, but also their similarity in shared haplotype frequencies and the lack of evidence for mitochondrial genetic structure found in this study.

The genetic analysis of salps is still in its infancy. This study brings the field forward through both genome and organism level insights on the Antarctic species *S. thompsoni*. Notably, the genome level insights were only uncovered by following up on signals of background variation within barcode sequencing profiles, suggesting that the investigation of similar signals in other animals might also be worthwhile. *S. thompsoni* provides fertile ground for further research on

mitochondrial DNA. Fundamental issues such as the number of mitochondrial isoforms and chromosomes still need to be established. Furthermore, despite the challenges associated with sequencing and assembling complicated mixtures of duplicated DNA, large scale sequencing is required to further elucidate the mechanisms responsible for the evolution of the *S. thompsoni* mitochondrial genome.

ACKNOWLEDGEMENTS

I would like to thank the JR26 participants and cruise leader Paul Rodhouse for collecting the *S. thompsoni* samples, Paul Giessler for assistance with the BAS biological archives, and Sílvia Pérez-Espona, Paul Seear and Jon Seger for their helpful comments. This research was supported by core funding from the Ecosystems group, British Antarctic Survey (Natural Environment Research Council).

REFERENCES

Arnheim N, Krystal M, Schmickel R, Wilson G, Ryder O, Zimmer E. 1980. Molecular evidence for genetic exchanges among ribosomal genes on non-homologous chromosomes in man and apes. *Proceedings of the National Academy of Sciences of the United States of America* **77**: 7323-7327.

Behura SK. 2007. Analysis of nuclear copies of mitochondrial sequences in honeybee (*Apis mellifera*) genome. *Molecular Biology and Evolution* **24**: 1492-1505.

Bensasson D, Zhang DX, Hartl DL, Hewitt GM. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology and Evolution* **16**: 314-321.

Birky CW Jr. 2001. The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annual Review of Genetics* **35**: 125-148.

Bryant D, Moulton V. 2004. Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution* **21**: 255-265.

Daponte MC, Capitanio FL, Esnal GB. 2001. A mechanism for swarming in the tunicate *Salpa thompsoni* (Foxton, 1961). *Antarctic Science* **13**: 240-245.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**: 47-50.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. 1994. DNA Primers for amplification of mitochondrial cytochrome C oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* **3**: 294-299.

Foxton P. 1961. *Salpa fusiformis* Cuvier and related species. *Discovery Reports* **32**: 1-32.

Frey JE, Frey B. 2004. Origin of intra-individual variation in PCR-amplified mitochondrial cytochrome oxidase I of *Thrips tabaci* (Thysanoptera: Thripidae): mitochondrial heteroplasmy or nuclear integration? *Hereditas* **140**: 92-98.

Gabriel W, Lynch M, Bürger R. 1993. Muller's ratchet and mutational meltdowns. *Evolution* **47**: 744-1757.

Gissi C, Iannelli F, Pesole G. 2008. Evolution of the mitochondrial genome of metazoa as exemplified by comparison of congeneric species. *Heredity* **101**: 301-320.

Goodall-Copestake WP, Pérez-Espona S, Clark MS, Murphy EJ, Seear PJ, Tarling GA. 2010. Swarms of diversity at the gene *cox1* in Antarctic krill. *Heredity* **104**: 513–518.

Goodall-Copestake WP, Tarling GA, Murphy EJ. 2012. On the comparison of population-level estimates of haplotype and nucleotide diversity: a case study using the gene *cox1* in animals. *Heredity* **109**: 50-56.

Goodall-Copestake WP. 2014. Morphological and molecular characterization of salps (*Thalia* spp.) from the Tristan da Cunha archipelago. *Journal of Plankton Research* **36**: 883-888.

Govindarajan AF, Bucklin A, Madin LP. 2011. A molecular phylogeny of the Thaliacea. *Journal of Plankton Research* **33**: 843-853.

Hazkani-Covo E, Zeller RM, Martin W. 2010. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS Genetics* **6**: e1000834.

Hebert PDN, Cywinska A, Ball SL, deWaard JR. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society B* **270**: 313-321.

Herd K, Barker SC, Shao R. 2012. High-level of heteroplasmy in the mitochondrial *cox1*-minichromosome of the human body louse, *Pediculus humanus*, and the human head louse, *Pediculus capitis*. *The Open Genomics Journal* **5**: 14-17.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**: 254-267.

Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. 2008. NCBI BLAST: a better web interface. *Nucleic Acids Research* **36** (Web Server issue): W5-W9.

Kajander OA, Rovio AT, Majamaa K, Poulton J, Spelbrink JN, Holt IJ, Karhunen PJ, Jacobs HT. 2000. Human mtDNA sublimons resemble rearranged mitochondrial genoms found in pathological states. *Human Molecular Genetics* **9**: 2821-2835.

Kayal E, Bentlage B, Collins AG, Kayal M, Pirro S, Lavrov DV. 2012. Evolution of linear mitochondrial genomes in medusozoan cnidarians. *Genome Biology and Evolution* **4**: 1-12.

Kmiec B, Woloszynska M, Janska H. 2006. Heteroplasmy as a common state of mitochondrial genetic information in plants and animals. *Current Genetics* **50**: 149-159.

Laslett D, Canbäck B. 2008. ARWEN, a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* **24**: 172-175.

Loeb VJ, Santora JA. 2012. Population dynamics of *Salpa thompsoni* near the Antarctic Peninsula: Growth rates and interannual variations in reproductive activity (1993–2009). *Progress in Oceanography* **96**: 93-107.

Mao M, Austin AD, Johnson NF, Dowton M. 2014. Coexistence of minicircular and a highly rearranged mtDNA molecule suggests that recombination shapes mitochondrial genome organization. *Molecular Biology and Evolution* **31**: 636-644.

Meyerhans A, Vatanian P, Wain-Hobson S. 1990. DNA recombination during PCR. *Nucleic Acids Research* **18**: 1687-1690.

Minin VN, Fang F, Dorman KS, Suchard MA. 2005. Dual multiple change-point model leads to more accurate recombination detection. *Bioinformatics* **21**: 3034-3042.

Nagarajan N, Pop M. 2013. Sequence assembly demystified. *Nature Reviews Genetics* **14**: 157-167.

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487-491.

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**: 5463-5467.

Singh TR, Tsagkogeorga G, Delsuc F, Blanquart S, Shenkar N, Loya Y, Douzery EJP. 2009. Tunicate mitogenomics and phylogenetics: peculiarities of the *Herdmania momus* mitochondrial genome and support for the new chordate phylogeny. *BMC Genomics* **10**: 534.

Thorpe SE, Heywood KJ, Stevens DP, Brandon MA. 2004. Tracking passive drifters in a high-resolution ocean model: implications for interannual variability of larval krill transport to South Georgia. *Deep Sea Research Part I: Oceanographic Research Papers* **51**: 909-920.

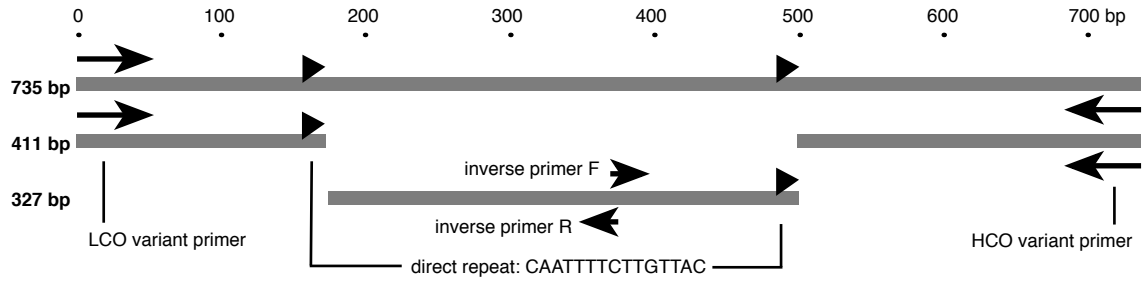
Venetis C, Theologidis I, Zouros E, Rodakis GC. 2006. No evidence for presence of maternal mitochondrial DNA in the sperm of *Mytilus galloprovincialis* males. *Proceedings of the Royal Society B* **273**: 2483-2489.

Wong LJ, Boles RG. 2005. Mitochondrial DNA analysis in clinical laboratory diagnostics. *Clinica Chimica Acta* **354**: 1–20.

Yokobori S, Fukuda N, Nakamura M, Aoyama T, Oshima T. 2004. Long-term conservation of six duplicated structural genes in cephalopod mitochondrial genomes. *Molecular Biology and Evolution* **21**: 2034-2046.

FIGURE 1

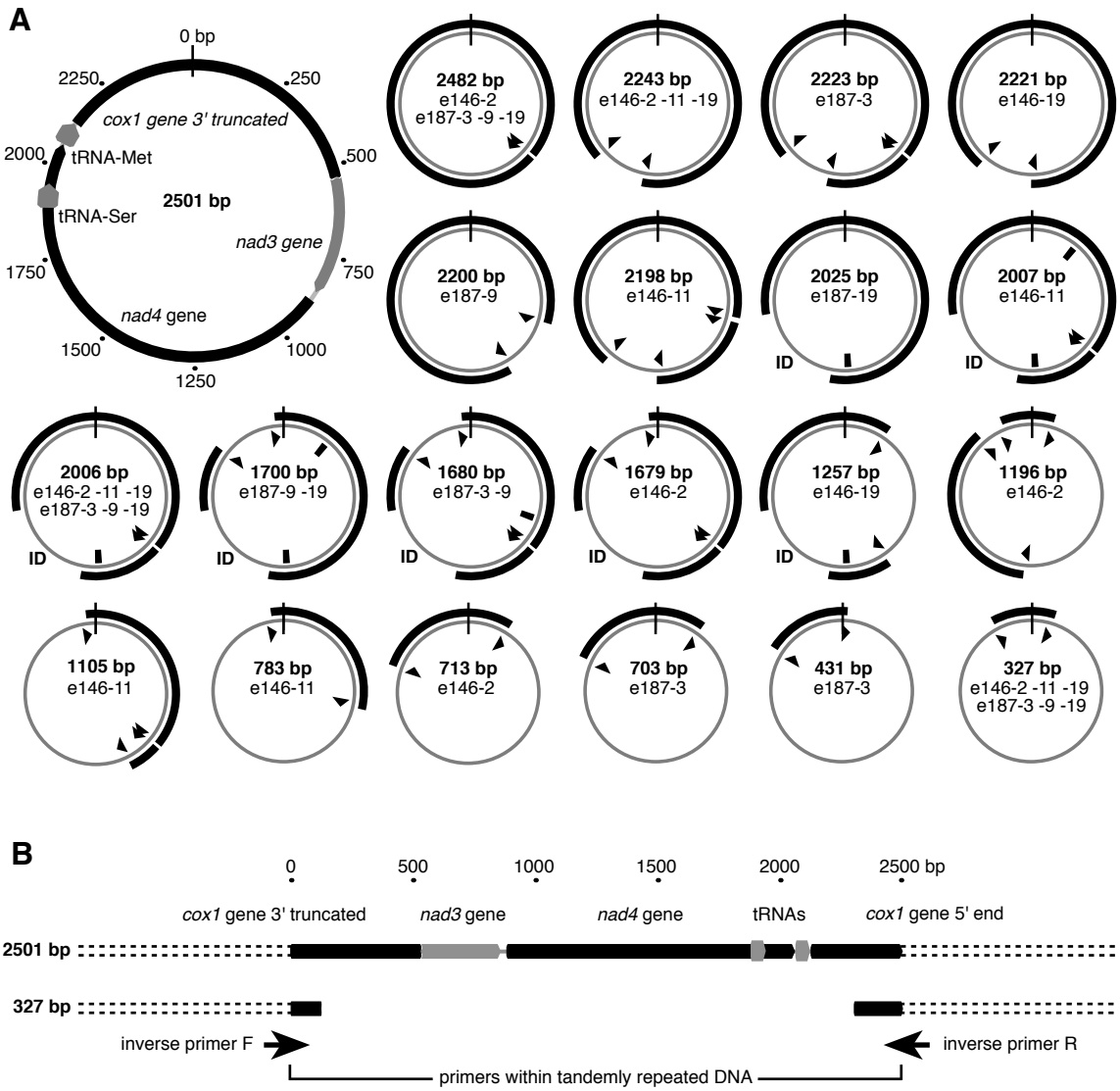
TITLE: Alignment of three *Salpa thompsoni* barcoding region amplicons



LEGEND: Amplicons from individual e187-9 obtained using the degenerate primers LCO variant and HCO variant (735 bp and 411 bp), and using the inverse PCR primers (327 bp). All amplicons were identical in sequence but not in length. The gap in the 411 bp amplicon exactly corresponds to the 327 bp amplicon - after accounting for three overlapping bases in the inverse primers - and is marked by the presence of a 15 bp direct repeat.

FIGURE 2

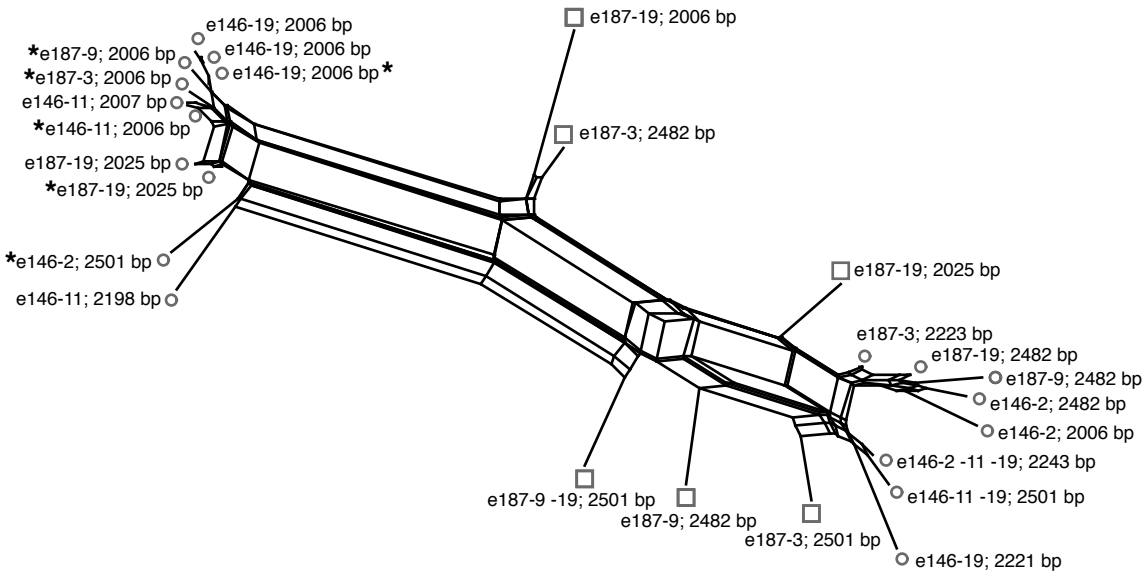
TITLE: Structural features of 21 different *Salpa thompsoni* inverse PCR amplicons



LEGEND: A) Circular molecule representation. The 2501 bp amplicon is annotated with bp markers and regions showing homology to mitochondrial genes and tRNAs. Position 0 (marked with a thin line) centres on the overlapping inverse PCR primers. Smaller amplicons are represented as re-arranged deletions of the 2501 bp amplicon with thick black lines indicating the extent of sequence retained. One gap corresponds to an inversion next to a deletion (labelled ID), the other gaps are simple deletions bordered by direct repeats (triangles). Single base insertions are shown as rectangles. The 2501 bp amplicon was recovered from all six individuals; the individuals containing other amplicon sizes are listed below the size labels. B) Linear molecule representation of the largest (2501 bp) and smallest (327 bp) amplicons nested within a longer region of tandemly repeated DNA (indicated by dashed lines).

FIGURE 3

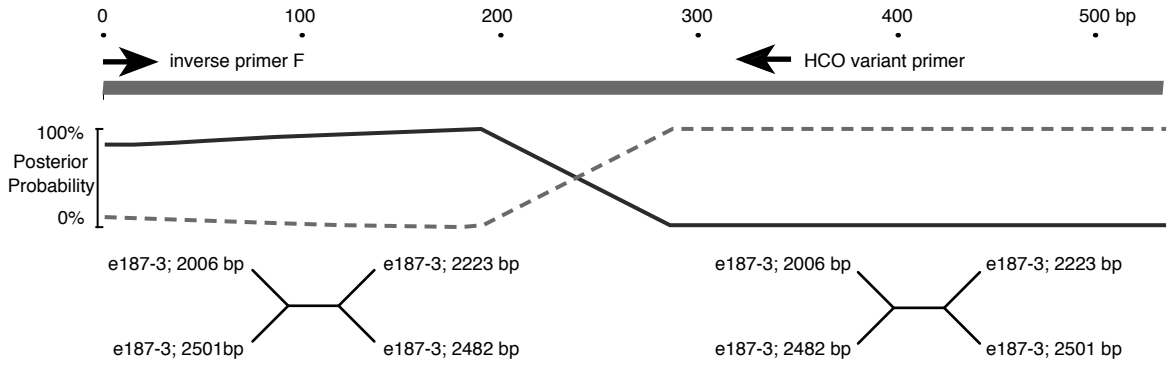
TITLE: Splits network for 29 *Salpa thompsoni* inverse PCR amplicons



LEGEND: Network generated using the Neighbor-Net method from uncorrected pairwise distances. Sequences containing obvious site-pattern conflicts within *cox1* are indicated by grey squares, the remainder are marked by grey circles. Names comprise the haul code (e146 or e187) followed by the sample number(s) and amplicon size; asterisks after names indicate a DNA barcode matching that from degenerate PCR on the same individual.

FIGURE 4

TITLE: Alternative phylogeny probabilities along a *Salpa thompsoni* DNA barcode



LEGEND: Phylogeny posterior probabilities were generated with the Geneious Dualbrothers plugin using four inverse PCR amplicon sequences derived from the same individual (e187-3). The black line corresponds to the topology on the left and dashed grey line corresponds to the topology on the right; primer regions are marked for orientation.

TABLE 1

TITLE: PCR primers used to amplify and sequence *Salpa thompsoni* DNA

Name	*Use*	*DNA sequence 5′ to 3′*
18S F	PCR and sequencing	CGGTAATTCCAGCTCCAAAAGTG
18S R	PCR and sequencing	GTTTCAGCTTTGCAACCATACTTC
LCO variant*	PCR	ACTGCCATCTTGGAGGACATTTTCTACAAAYCATAARVATATYRS
HCO variant*	PCR	CCCAGGATTATGGAACGGTAAACTTCTGGRTGNCYGAARAAYCA
TAILED*	PCR tests and sequencing	ACTGCCATCTTGGAGGAC
PRIMER*	PCR tests and sequencing	CCCAGGATTATGGAACGG
Inverse F	PCR and sequencing	ATGAACAATTTACCCACCCCTGAG
Inverse R	PCR and sequencing	CATCTACATGAAGGAGCTACAGATG
IntSeq1 F	Sequencing	CTATATTTGAGTGCAGATTTGAATCG
IntSeq1 R	Sequencing	CGATTCAAATCTGCACTCAAATATAG
IntSeq2 F	Sequencing	AGGAAAAGGTAAGTCTATAAGCAG
IntSeq2 R	Sequencing	CTGCTTATAGACTTACCTTTTCCT
TAILEDinv F	PCR tests	ACTGCCATCTTGGAGGACATGAACAATTTACCCACCCCTGAG
18-MERinv F	PCR tests	ACACAGAGGCTTTCTGCGATGAACAATTTACCCACCCCTGAG
18-MERinv R	PCR tests	ACACAGAGGCTTTCTGCGCATCTACATGAAGGAGCTACAGATG
PRIMERinv R	PCR tests	CCCAGGATTATGGAACGGCATCTACATGAAGGAGCTACAGATG
18-MER	PCR tests	ACACAGAGGCTTTCTGCG

LEGEND: All of the primers were specifically designed for this study except those marked * from Goodall-Copestake (2014)

TABLE 2

TITLE: Barcode (*cox1*) diversity in *Salpa thompsoni* oozooid samples

Sampling region, code	N	*Haplotypes*	*h* ±*s.d.*	*π* ±*s.d.*	^{456} *h* ±*s.d.*	^{456} *π* ±*s.d.*
South Georgia, e146	24	6	0.442 +/–0.124	0.00076 +/–0.00075	0.442 +/–0.124	0.00108 +/–0.00107
Antarctic Peninsula, e187	24	8	0.609 +/–0.112	0.00135 +/–0.00110	0.496 +/–0.119	0.00137 +/–0.00124

LEGEND: Abbreviations: *h,* haplotype diversity; *π,* nucleotide diversity; N, sample size; s.d., standard deviation; ^{456} diversity metrics recalculated from 456 bp of barcode (*cox1*) DNA.