Novel diversity in mitochondrial genomes of deep-sea Pennatulacea (Cnidaria: Anthozoa: Octocorallia)

Raissa I. Hogan, Kevin Hopkins, Andrew J. Wheeler, A. Louise Allcock & Chris Yesson

View supplementary material

Published online: 18 Jul 2019.

Submit your article to this journal

Article views: 345

View related articles

View Crossmark data

RESEARCH ARTICLE

Novel diversity in mitochondrial genomes of deep-sea Pennatulacea (Cnidaria: Anthozoa: Octocorallia)

Raissa I. Hogan[a] , Kevin Hopkins[b], Andrew J. Wheeler[c] , A. Louise Allcock[a] and Chris Yesson[b]

[a]Department of Zoology, Ryan Institute, National University of Ireland, Galway, Ireland; [b]Institute of Zoology, Zoological Society of London, Regent's Park, London, UK; [c]School of Biological, Earth and Environmental Sciences/iCRAG/ERI, University College Cork, Cork, Ireland

ABSTRACT

We present the first documented complete mitogenomes of deep-sea Pennatulacea, representing nine genera and eight families. These include one species each of the deep-sea genera *Funiculina*, *Halipteris*, *Protoptilum* and *Distichoptilum*, four species each of *Umbellula* and *Pennatula*, three species of *Kophobelemnon* and two species of *Anthoptilum*, as well as one species of the epi- and mesobenthic genus *Virgularia*. Seventeen circular genomes ranged from 18,513 bp (*Halipteris* cf. *finmarchica*) to 19,171 bp (*Distichoptilum gracile*) and contained all genes standard to octocoral mitochondrial genomes (14 protein-coding genes, two ribosomal RNA genes and one transfer RNA). We found at least three different gene orders in Pennatulacea: the ancestral gene order, the gene order found in bamboo corals (Family Isididae), and a novel gene order. The mitogenome of one species of *Umbellula* has a bipartite genome (~13 kbp and ~5 kbp), with good evidence that both parts are circular.

Introduction

Pennatulacea was described as one of the strangest wonders when first reported in the early part of the seventeenth century (Kerr, 1824, in Williams 2011). The 'wonders' referred to its very distinctive and highly specialized morphological characteristics. Pennatulaceans are colonial marine anthozoans in the subclass Octocorallia. They differ from all other octocorals by having from three to five differentiated and specialized polyp types (Williams et al. 2012). The primary polyp (oozooid) is modified into the rachis (the distal region where the secondary polyps rise) and peduncle (the proximal region that anchors the colony in mostly soft substrates [Williams and Alderslade 2011]).

Sea pens comprise approximately 200 species in 37 genera and 14 families (Pérez et al. 2016), with almost one-third of genera being classified as deep-water (maximum depth >2000 m) (Williams 2011). They have a broad distribution from tropical zones to the poles and from shallow waters to depths reaching more than 6000 m (Williams 2011). Yesson et al. (2012) conducted global habitat suitability analyses for cold-water octocorals and predicted that the pennatulacean suborder Sessiliflorae had the widest potential habitat range. In the deep sea, sea pens can form single or multi-species assemblages, commonly known as 'coral-gardens' or 'sea pen fields', where they play a pivotal role as engineer species, forming a complex three-dimensional habitat providing shelter, nursery and food for a range of species, including commercial fish (Baillon et al. 2012, 2014; De Clippele et al. 2015). To date, the evolutionary history of sea pens still remains unclear, and molecular data are scarce (Berntson et al. 2001; McFadden et al. 2006; Dolan et al. 2013; Kushida and Reimer 2018), with only two studies addressing phylogenetic relationships within sea pen groups (Dolan et al. 2013; Kushida and Reimer 2018).

Complete mitochondrial genomes have proven to be a useful tool in recent studies as they provide a substantial quantity and diversity of DNA data (Lavrov 2007; Kayal et al. 2012). They have been effective in phylogenetic and molecular evolutionary studies at different taxonomic levels (Kayal and Lavrov 2008; Brockman and McFadden 2012; Figueroa and Baco 2013; Kayal et al. 2013; Lavrov and Pett 2016).

The first octocoral mitochondrial genome published was from a Pennatulacea (*Renilla koellikeri*) (Beagley et al. 1995) followed by the octocoral *Sarcophyton glaucum* (Beaton et al. 1998; Pont-Kingdon et al. 1998). Their single circular mitochondrial structures had the same gene order and were compared to the other metazoans. Their general composition was described in detail, including a reduction in the number of transfer ribonucleic acid genes (*tRNA*) and the presence of a putative mismatch repair gene, *mtMutS*, a molecular synapomorphy in Octocorallia.

Currently, there are 55 octocorallian mitochondrial genomes deposited in GenBank, which represent less than 2% of all octocoral species. To date, six different gene orders

⊕ Supplemental data for this article can be accessed here.

have been described (Beagley et al. 1995; Brugler and France 2008; Uda et al. 2011; Brockman and McFadden 2012; Pante et al. 2013). The most commonly encountered gene order was first described in *Renilla koellikeri* (Beagley et al. 1995) and subsequently *Sarcophyton glaucum* (Beaton et al. 1998), is present in all major clades of octocorals (McFadden et al. 2006; Figueroa and Baco 2015), and is considered the possible octocoral 'ancestral gene order'. It is referred to as gene order A (Brockman and McFadden 2012; Park et al. 2012). Subsequent novel gene orders have been named in order of discovery, and they have been related to a specific taxon/clade: gene order B, bamboo coral BAL208-1 (Family Isididae) (Brugler and France 2008); gene order C and D, *Paracorallium japonicum* and *Corallium konojoi*, respectively (Family Corallidae) (Uda et al. 2011); gene order E, *Paraminabea aldersladei* (Family Alcyoniidae) (Brockman and McFadden 2012); and gene order F, *Isidoides armata* (Suborder Calcaxonia: Family *incertae sedis*) (Pante et al. 2013). Gene orders A to E are known from complete mitochondrial genome sequences (Beagley et al. 1995; Brugler and France 2008; Uda et al. 2011; Brockman and McFadden 2012), while gene order F was inferred from gene junction screening (Pante et al. 2013).

In five of the six gene orders described to date, four blocks of genes are conserved (**Block 1**: cox1-rns-nad1-cob; **block 2:** nad6-nad3-nad4; **block 3:** mtMutS-rnl-nad2-nad5-nad4; **block 4:** trnM-cox3-atp6-atp8-cox2) (Brockman and McFadden 2012). Gene junction sequencing did not elucidate the full arrangement of gene order F, but it does appear that not all blocks are conserved since cox1 and cob are widely separated in the genome. The characteristics of the five first known gene orders led Brockman and McFadden (2012) to suggest the arrangements could have arisen through inversions and translocations of these large blocks.

Thus, our current understanding is that a 'typical' octocoral mitochondrial genome is composed of a single circular chromosome, fourteen protein-coding genes (atp6, atp8, cox1-3, cob, nad1-6, nad4L and mtMutS), two ribosomal RNA subunits (rns or 12S and rnl or 16S) and one transfer RNA (trnM), ranging from 17,358 bp (*Muricea purpurea*; Uda et al. 2013) to 20,246 bp (*Calicogorgia granulosa*; Park et al. 2011).

Despite the significance of the order Pennatulacea, little attention has been paid to its molecular biology and evolution. Twenty-three years after the first sea pen mitochondrial genome description, the complete mitochondrial genome has been sequenced for only two other species (*Renilla muelleri* and *Stylatula elongata*) (Kayal et al. 2013), neither of which is present in deep water. In addition, both were sequenced as part of a higher systematic level study (Kayal et al. 2013), and no descriptive analyses of mitochondrial genome structure were published.

That three published sea pen mitochondrial genomes have the same gene order (ancestral gene order), with general characteristics consistent with other octocorals, raises the possibility of a conserved mitochondrial genome in sea pens.

This study seeks to address some of these research gaps and contribute to a deeper understanding of Pennatulacea molecular evolution. We sequenced 18 complete mitochondrial genomes, including nine genera and eight families and encompassing almost all genera present in the Northeast Atlantic and deep water globally, with the aim of extending our understanding of mitochondrial genome evolution in sea pens, and consequently in octocorals.

Materials and methods

Sampling and morphological species identification

Pennatulaceans were collected from RV *Celtic Explorer* using the remotely operated vehicle (ROV) *Holland I* between 2013 to 2016, and from collaborators from surveys in UK, Ireland, Greenland and Bay of Biscay (Table 1). After extensive ROV surveys in the Northeast Atlantic Ocean, we selected eighteen samples we considered would represent most of the morphotypes present in this region. Specimens were identified by the authors with reference to *in situ* and *ex situ* photos and by examining morphological characters under a stereo microscope, with reference to classical and current

Table 1. Collection information, museum vouchers, and GenBank Accession numbers for samples used herein.

Species	Depth (m)	Location	Accession Number	Genbank Number
Anthoptilum grandiflorum	1,150	Greenland	NMS.Z.2019.25.16	MK91965
Anthoptilum sp. 1	3,148	Whittard Canyon	NMS.Z.2019.25.1	MK919656
Distichoptilum gracile	2,350	Whittard Canyon	NMS.Z.2019.25.2	MK919657
Funiculina quadrangularis	15-40	Little Loch Broom, Scotland	NMS.Z.2019.25.17	MK919658
Halipteris cf. *finmarchica*	1,891	Whittard Canyon	NMS.Z.2019.25.3	MK919659
Kophobelemnon sp. 1	1,113	Whittard Canyon	NMS.Z.2019.25.4	MK919660
Kophobelemnon sp. 3	1,606	Whittard Canyon	NMS.Z.2019.25.5	MK919661
Kophobelemnon sp. 4	760	Whittard Canyon	NMS.Z.2019.25.6	MK919662
Pennatula aculeata	1,726	Whittard Canyon	NMS.Z.2019.25.7	MK919663
Pennatula cf. *aculeata*	1,000	Bay of Biscay		MK919664
Pennatula cf. *inflata*	1,830	Whittard Canyon	NMS.Z.2019.25.8	MK919665
Pennatula grandis	1,338	Whittard Canyon	NMS.Z.2019.25.9	MK919666
Protoptilum carpenteri	965	Whittard Canyon	NMS.Z.2019.25.10	MK919667
Umbellula huxleyi	1,772	Whittard Canyon	NMS.Z.2019.25.11	MK919668
Umbellula sp. 1	2,322	Whittard Canyon	NMS.Z.2019.25.12	MK919669
Umbellula sp. 2	1,708	Porcupine Bank, Ireland	NMS.Z.2019.25.13	MK919670
Umbellula sp. 3	4,173	Whittard Canyon	NMS.Z.2019.25.14	MK919672 (13kb) MK919671 (5 kb)
Virgularia mirabilis	7	Galway Bay, Ireland	NMS.Z.2019.25.15	MK919673

taxonomical literature. We were conservative in our species attributions, preferring to identify only to genus when appropriate, numbering multiple putative species from the same genus as sp. 1, sp. 2 etc. Voucher material for the 18 specimens is deposited at National Museums Scotland, Edinburgh, UK (Table 1).

Mitochondrial genome sequencing and assembly

Genomic DNA was extracted from tissue using the NucleoSpin® Tissue Kit (Macherey-Nagel, Düren, Germany) following the manufacturer's instructions. DNA yield was assessed using a Qubit 2.0 Fluorometer (Life Technologies). Samples with sufficient yield (30 ng/μl – 38 ng/μl) were selected for sequencing. Genomic DNA was fragmented using a Covaris M220 Focused-ultrasonicator™.

Indexed libraries were prepared using a TruSeq DNA PCR-Free Library Preparation Kit (Illumina Inc., San Diego, CA) and sequenced on an Illumina MiSeq (MiSeq Reagent Kit v3 600-cycle) at the Institute of Zoology's NGS lab. Following qPCR library quantification, two samples (*Kophobelemnon* sp. 4 and *Umbellula huxleyi*) did not contain enough library to proceed so they were amplified using the enrichment protocol from the TruSeq Nano protocol.

Prior to quality control editing and genome assembly, we combined the paired reads using Geneious set paired reads option (Kearse et al. 2012). The quality of the reads was assessed with FastQC (Andrews 2010). Illumina TruSeq adapters and low-quality reads (Q < 30) were trimmed using first cutadapt 1.9.1 (Martin 2011) and subsequently fastx-trimmer in FASTX-Toolkit 0.0.13 (Gordon and Hannon, 2010).

Mitochondrial genomes were constructed using *de novo* assembly in MITObim (Hahn et al. 2013) using partial mitochondrial sequences (*mtMutS* or *cox1*) from each species as initial seed references. *Anthoptilum* sp. 1 failed *de novo* assembly, and instead the MIRA mapping assembly was used in MITObim (Hahn et al. 2013) using the whole mitochondrial genome of *Anthoptilum grandiflorum* as a reference.

Sanger sequencing of gene boundaries

To confirm an unusual gene arrangement in *Umbellula* sp. 3, we also constructed the genome using NOVOPlasty 3.1 (Dierckxsens et al., 2017) and obtained an identical arrangement. Additionally, certain gene boundaries were amplified using a range of known and new primers (Table 2) in standard polymerase chain reactions (PCRs), and sequenced using Sanger methods. The PCR protocol comprised an initial denaturing step of 94 °C for 2 min; 35 cycles at 94 °C for 30 s, 55 °C for 60 s and 72 °C for 120 s, followed by a 10 min extension step at 72 °C. PCR solutions contained 12 μl GoTaq G2 Green Mmix (Promega), 9 μl sterile H_2O, 0.5 μl forward primer [10 μM], 0.5 μl reverse primer [10 μM], and 2.5 μl of DNA template. DNA was purified using Pure Link, PCR Purification Kit, Invitrogen (Life Technologies) following the manufacturer's instructions. Clean PCR products were sent to GATC Biotech, Germany, for sequencing. Details of regions amplified are given after the description of the genome for clarity.

Mitochondrial genome annotation and gene order

Reads were mapped to the assembled sequences using the Geneious Read Mapper (Kearse et al. 2012) to check the coverage quality of the entire genome and confirm the gene orders.

Genes were identified using Mitos2 (Bernt et al. 2013), with NCBI RefSeq 81 Metazoa database reference and genetic code 4, which includes Coelenterata. Annotations were edited in Geneious 11.1.5 (Kearse et al. 2012) with regard to the widest open reading frame (ORF) using Cnidarian start (ATG) and stop (TAG, TAA) codons to determine the protein-gene boundaries. Transfer RNAs (*tRNA*) were identified using MITFI (Jühling et al. 2012) and rRNAs using a structure-based covariant model, both implemented in Mitos2.

Alternative start and stop codons (e.g. TGA, ATA, AGA, AGG) have been proposed for non-cnidarian Metazoa and alternative codes have also been suggested for some octocorals (Brockman and McFadden 2012; Kayal et al. 2012). However, since no studies have validated the latter, for consistency, we followed the standard cnidarian genetic code. Once the genes were annotated, the intergenic regions and overlapping genes of complete mitochondrial genomes were determined manually.

Mitochondrial gene order rearrangements were compared using CREx (Bernt et al. 2007). It uses pair-wise comparisons to identify possible evolutionary scenarios between two gene orders considering transpositions, reverse transpositions, reversals, and tandem-duplication-random-loss (tdrl) events.

Table 2. PCR primers for Sanger sequencing across gene junctions. *Modified slightly from cited publication.

Primer	Binding Region	Sequence	Author
13F	*mtMutS*	CTATTTTAGGYTGGAAGAGA	Park et al. (2010)
13R	*rnl*	CTGTTTCCAAGCCTACTT	Park et al. (2010)
ND42599F	*nad4L*	GCCATTATGGTTAACTATTAC	France and Hoover (2002)
Mut3458R	*mtMutS*	TSGAGCAAAAGCCACTCC	Sánchez et al. (2003)
16S647F	*rnl*	ACACAGCTCGGTTTCTATCTACAA*	McFadden et al. (2006)
msh2806R	*mtMutS*	TAACCCAGCTTGGGAGTGTGC*	Brugler & France (2008)
COII8068F	*cox2*	CCATAACAGGACTAGCAGCATC	McFadden et al. (2006)
COI8325R	*cox1*	TCCTTATGATTAGTAGAAAA	McFadden et al. (2006)
cobPenF	*cob*	TTAGGAGCCAACCCGGTAGA	this study
nad6PenR	*nad6*	TAGCTCCTACCATGGCGACT	this study
ND4gen4974F	*nad4*	TAGGCTTATTTACTCATACGAT*	Brugler and France (2008)
CO3bam5657F	*cox3*	GCTGCTAGTTGGTATTGGCAT	Brugler and France (2008)

Mitochondrial genome composition

Nucleotide composition and GC content were calculated in Geneious 11.1.5 (Kearse et al. 2012). Strand-specific compositional asymmetry was calculated following the skew formula of Perna and Kocher (1995). The accumulative GC skews were visualized using GraphDNA (Thomas et al. 2007). Positive values of GC skew and AT skew indicate that the strand being analyzed has more Gs than Cs, and more As than Ts, respectively. The leading strand generally contains more guanine and thymine, therefore the leading strand is expected to have a positive GC skew and a negative AT skew, versus negative GC skew and positive AT skew in the lagging strand.

The DNA-Walk method (Lobry 1996) in GraphDNA was used to identify origins of replications in the heavy (oriH) and light (oriL) strands considering that these regions could be associated with abrupt reversals in base composition bias.

Tandem repeats and inverted repeats were sought across the whole genome using Tandem Repeats Finder (Benson 1999) and EMBOSS einverted (Rice et al. 2000). Presence of stable stem-loops was investigated using MFold (Zuker 2003) in areas where we identified an abrupt change in base composition bias. These features may also be associated with the origin of replication.

Results

Genome organization

We assembled mitochondrial genomes of 18 species/morphotypes of Northeast Atlantic deep-sea Pennatulacea, representing nine genera and eight families (Table 3). We recovered seventeen circular genomes that ranged from 18,513 (Halipteris cf. finmarchica) to 19,171 bp (Distichoptilum gracile). The mitochondrial genome of Umbellula sp. 3 was recovered as two separate circular chromosomes (∼13 kbp and ∼5 kbp) with a total length of 18,517 bp. Using de novo assembly, we recovered only 11 kbp of Anthoptilum sp. 1. Using the newly recovered mitochondrial genome of Anthoptilum grandiflorum as a backbone, and mapping reads against it, we managed to obtain a circular sequence of

18,842 bp of Anthoptilum sp. 1 including all the genes that have been described for Pennatulacea species.

All eighteen mitochondrial genomes have fourteen protein-coding genes (atp6, atp8, cox1-3, cob, nad1-6, nad4L and mtMutS), two ribosomal RNA subunits (rns or 12S and rnl or 16S) and one transfer RNA (trnM) (Tables 4 and 5). Most of the genes are separated by intergenic regions (IGR). The largest IGRs are found between nad4 and trnM in Distichoptilum gracile, Kophobelemnon sp. 3 and Kophobelemnon sp. 4, and extend for 566 bp, 523 bp and 504 bp respectively. While the IGR between cob and nad6 is the second largest, extending in Pennatula cf. inflata for 400 bp, it is also the smallest, extending in Umbellula huxleyi for 1 bp. The IGR between nad3-nad4L extends for only 3 bp in all species (Tables 4 and 5).

Overlapping genes are present in all genomes, with two to five pairs of genes overlapping per genome. These include nad1-cob (94–119 bp), nad5-nad4 (17–33 bp), nad6-nad3 (14 bp), nad5-nad2 (13 bp), rnl-nad2 (68 bp) and mtMutS-rnl (28 bp). In the two Umbellula species that have different gene orders, two pairs of genes overlapped, cob-nad6 (8 bp) and nad2-nad5 (13 bp) (Tables 4 and 5).

Gene order

We found at least three different genome arrangements in Pennatulacea (Figure 1). Fourteen species have the same arrangement as the octocoral ancestral gene order (gene order A; Beagley et al. 1995). Anthoptilum species have the same gene order as the bamboo corals Isididae sp. (EF622534) and Acanella eburnea (EF672731) (gene order B; Brugler and France 2008). Umbellula sp. 1 has a novel gene order (gene order G), which is the seventh reported octocoral mitochondrial genome arrangement.

Results for one species of Umbellula (Umbellula sp. 3) indicate a bipartite mitochondrial genome. Two isolated contigs (∼13 kbp and ∼5 kbp) were recovered from the de novo assembly. The larger contig has eleven protein-coding genes (atp6, atp8, cox1-3, cob, nad1-2, nad4-6), one ribosomal RNA gene (rns) and one transfer RNA (trnM), and the smaller has

Table 3. Pennatulacea samples and NGS data details used in this study.

Species	Reads	Clean	MtDNA (bp)	Coverage mean (min–max)	Gene Order
Anthoptilum grandiflorum	3,981,221	3,416,102	19,391	73 (39–114)	B
Anthoptilum sp. 1	2,656,633	2,287,755	18,850	16 (5–32)	B
Distichoptilum gracile	4,675,062	3,891,766	19,171	63 (42–88)	A
Funiculina quadrangularis	5,665,947	4,786,607	18,906	270 (191–351)	A
Halipteris cf. finmarchica	2,684,606	2,320,553	18,513	25 (5–44)	A
Kophobelemnon sp. 1	3,064,336	2,542,074	18,883	40 (20–61)	A
Kophobelemnon sp. 3	2,460,329	2,084,735	19,109	33 (12–57)	A
Kophobelemnon sp. 4	2,222,497	1,847,321	19,130	36 (16–63)	A
Pennatula aculeata	3,475,982	3,013,797	18,715	75 (34–129)	A
Pennatula cf. aculeata	3,062,179	2,550,313	18,715	128 (82–170)	A
Pennatula cf. inflata	5,033,535	4,199,539	19,127	111 (74–153)	A
Pennatula grandis	3,472,717	2,991,151	18,973	122 (69–179)	A
Protoptilum carpenteri	736,848	615,240	18,729	21 (6–38)	A
Umbellula huxleyi	1,973,389	1,656,462	18,927	45 (19–81)	A
Umbellula sp. 1	3,745,361	3,137,577	18,714	68 (40–94)	G (new)
Umbellula sp. 2	4,205,005	3,622,094	18,767	32 (12–54)	A
Umbellula sp. 3 _ 13k	1,209,061	1,011,587	12,938	41 (22–77)	Contig i
Umbellula sp. 3 _ 5k	1,209,061	1,011,587	5,579	22 (10–40)	Contig ii
Virgularia mirabilis	1,863,469	1,617,766	18,770	38 (9–74)	A

Table 4. Pennatulacea mitochondrial gene lengths (base pairs), intergenic regions (positive values) and overlapping genes (negative values) (base pairs) in specimens in this study with gene order A. Intergenic regions are represented in italics.

Genes	Distichoptilum gracile	Funiculina quadrangularis	Halipteris cf. finmarchica	Kophobelemnon sp. 1	Kophobelemnon sp. 3	Kophobelemnon sp. 4	Pennatula aculeata	Pennatula cf. aculeata
cox2-cox1	**124**	**124**	**61**	**124**	**124**	**124**	**124**	**124**
cox1	1566	1566	1563	1566	1566	1566	1566	1566
cox1-rns	*37*	*37*	*27*	*37*	*37*	*37*	*37*	*37*
rns	1057	1057	1065	1057	1057	1058	1057	1057
rns-nad1	*51*	*51*	*45*	*52*	*52*	*63*	*33*	*33*
nad1	981	981	1068	981	981	981	1092	1092
nad1-cob	*−94*	*−94*	*−95*	*−94*	*−94*	*−94*	*−119*	*−119*
cob	1266	1266	1188	1266	1266	1266	1185	1185
cob-nad6	*18*	*18*	*26*	*18*	*18*	*18*	*18*	*18*
nad6	555	555	546	555	555	555	555	555
nad6-nad3	*64*	*64*	*40*	*64*	*64*	*64*	*23*	*23*
nad3	354	354	354	354	354	354	354	354
nad3-nad4L	*3*	*3*	*3*	*3*	*3*	*3*	*3*	*3*
nad4L	294	294	294	294	294	294	294	294
nad4L-mtMutS	*13*	*13*	*13*	*13*	*13*	*13*	*13*	*13*
mtMutS	2961	2955	2940	2955	2955	2955	2994	2994
mtMutS-rnl	*154*	*154*	*−28*	*154*	*154*	*154*	*154*	*154*
rnl	1788	1788	2066	1766	1765	1783	1788	1788
rnl-nad2	*64*	*37*	*−68*	*37*	*37*	*37*	*64*	*64*
nad2	1356	1380	1365	1380	1380	1380	1356	1356
nad2-nad5	*−13*	*−13*	*−13*	*−13*	*−13*	*−13*	*−13*	*−13*
nad5	1818	1818	1815	1818	1818	1818	1818	1818
nad5-nad4	*−33*	*−33*	*−17*	*−28*	*−28*	*−28*	*−33*	*−33*
nad4	1449	1449	1449	1449	1449	1449	1449	1449
nad4-trnM	*566*	*345*	*63*	*296*	*523*	*504*	*117*	*117*
trnM	71	71	71	71	71	71	71	71
trnM-cox3	*50*	*64*	*48*	*64*	*64*	*64*	*64*	*64*
cox3	786	786	786	786	786	786	786	786
cox3-atp6	*57*	*45*	*44*	*57*	*57*	*57*	*57*	*57*
atp6	708	708	708	708	708	708	708	708
atp6-atp8	*20*	*17*	*20*	*33*	*33*	*33*	*20*	*20*
atp8	216	216	216	216	216	216	216	216
atp8-cox2	*38*	*28*	*30*	*28*	*28*	*38*	*38*	*38*
cox2	762	762	789	762	762	762	762	762
Total IGR (bP)	1259	1000	420	980	1207	1209	765	765
% IGR/mtgen	6.57	5.29	2.27	5.19	6.32	6.32	4.09	4.09
#ol genes	3	3	5	3	3	3	3	3
% pcg	79	80	81	80	79	79	81	81

three protein-coding genes (*mtMutS, nad3, nad4L*) and one ribosomal RNA gene (*rnl*).

To investigate alternative gene orders for this species, we attempted read mapping against each of the six gene arrangements for which whole mitochondrial genome sequences were available (Online Resource 1 (ESM 1)). Mapping for the bipartite arrangement recovered relatively high and continuous coverage of each contig indicative of two circular contigs (Online Resource 2 (ESM 2)) (*Umbellula* sp. 3, Coverage: 13 kbp contig - min 22x/max 77x/average 1x; 5 kbp contig - min 12x/max 18x/average 14.5x). In contrast, read mapping to all other known gene orders produced zero coverage at the gene boundaries not found in the bipartite arrangement (Online Resource 1 (ESM 1)). We also tried seeding MITObim with each gene in turn, and always recovered the same two contigs. This supports the bipartite gene arrangement and rejects all other known gene arrangements.

We used PCR to amplify across gene block and other gene boundaries (see Figure 1 where gene blocks 1–4 are colored red, blue, yellow and green). In the small contig, we successfully amplified and sequenced across the gene junction of *mtMutS* and *rnl* using primers 13F and 13R (GenBank Accession No: MK992494), across the *nad4L* and *mtMutS* gene junction using primers ND42599F and Mut3458R

(GenBank Accession No: MK992495), and a region extending from *rnl* across *nad3*, *nad4L*, to the start of *mtMutS* using the primers 16S647F and msh2806R (GenBank Accession No: MK992492), confirming the circularity of the smaller contig. In the large contig, we successfully amplified and sequenced across the gene junction of *cox2* and *cox1* using primers COII8068F and COI8325R (GenBank Accession No: MK992496), across the gene junction of *cob* and *nad6* using primers cobPenF and nad6PenR (GenBank Accession No: MK992497), and from *nad4* across *trnM* to *cox3* using primers ND4gen4974F and CO3bam5657F (GenBank Accession No: MK992498), suggesting this contig is also likely circular. As expected, a PCR using the primers CO3bam5657F and 13 R produced no amplification product, indicating that *cox3* and *rnl* (where these primers bind) are not close/adjacent in *Umbellula* sp. 3 (compare with the arrangement of *Umbellula* sp. 1 [Figure 1] where we would expect to obtain an amplification product, and in fact did [data not shown]).

The two parts of this bipartite genome could be derived from either gene order A or gene order G (the new gene order) with the *mtMutS-nad4L-nad3-rnl* block extracted into a separate loop (Figure 1). Since *Umbellula* sp. 3 has genes and intergenic regions with sizes and compositions more similar to other species with gene order G (Table 5) we believe gene order G represents the mostly likely origin.

Table 4. Continued.

Genes	Pennatula cf. inflata	Pennatula grandis	Protoptilum carpenteri	Umbellula. huxleyi	Umbellula sp. 2	Virgularia mirabilis
cox2-cox1	**124**	**124**	**124**	**73**	**101**	**102**
cox1	1566	1566	1566	1566	1566	1566
cox1-rns	*37*	*37*	*37*	*49*	*37*	*37*
rns	1057	1057	1057	1062	1059	1056
rns-nad1	*33*	*33*	*51*	*65*	*51*	*51*
nad1	999	999	981	963	981	981
nad1-cob	*−94*	*−94*	*−94*	*15*	*−94*	*−94*
cob	1275	1275	1266	1593	1266	1266
cob-nad6	*400*	*246*	*18*	*1*	*18*	*18*
nad6	555	555	555	573	555	555
nad6-nad3	*64*	*64*	*64*	*−14*	*64*	*64*
nad3	354	354	354	381	354	354
nad3-nad4L	*3*	*3*	*3*	*3*	*3*	*3*
nad4L	294	294	294	294	294	294
nad4L-mtMutS	*13*	*13*	*13*	*6*	*13*	*13*
mtMutS	2967	2967	2970	2925	2961	2991
mtMutS-rnl	*155*	*155*	*154*	*154*	*157*	*154*
rnl	1788	1788	1765	1760	1769	1799
rnl-nad2	*65*	*65*	*37*	*41*	*65*	*80*
nad2	1356	1356	1383	1374	1356	1356
nad2-nad5	*−13*	*−13*	*−13*	*−13*	*−13*	*−13*
nad5	1818	1818	1818	1815	1818	1815
nad5-nad4	*−33*	*−33*	*−33*	*−20*	*−33*	*−28*
nad4	1449	1449	1449	1449	1449	1449
nad4-trnM	*123*	*123*	*123*	*63*	*204*	*118*
trnM	71	71	71	71	71	71
trnM-cox3	*50*	*50*	*64*	*68*	*50*	*64*
cox3	786	786	786	828	786	786
cox3-atp6	*57*	*57*	*57*	*5*	*50*	*57*
atp6	708	708	708	708	708	708
atp6-atp8	*20*	*20*	*20*	*17*	*20*	*23*
atp8	216	216	216	216	216	216
atp8-cox2	*38*	*38*	*38*	*15*	*38*	*42*
cox2	762	762	762	783	762	762
Total IGR (bp)	1182	1028	803	546	871	826
% IGR/mtgen	6.18	5.42	4.29	2.88	4.64	4.40
#ol genes	3	3	3	3	3	3
% pcg	79	80	81	82	80	80

IGR: intergenic region; #ol genes: number of overlapping genes; pcg: protein-coding genes. OriH and OriL indicated by bold and underline, respectively.
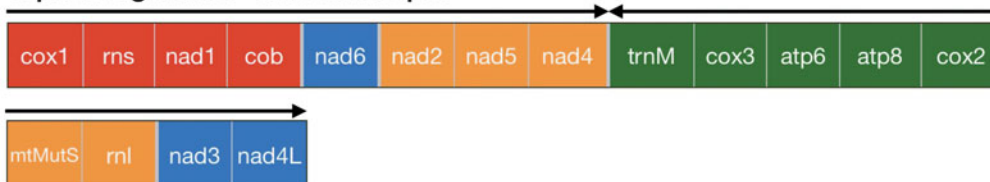

Figure 1 Pennatulacea mitochondrial genome arrangements (linearized). Gene order A (octocoral ancestral gene order): 14 pennatulacean species of our study (Table 1). Gene order B: *Anthoptilum* spp. New gene order (G): *Umbellula* sp. 1. Bipartite mitochondrial genome: *Umbellula* sp. 3. Different colors correspond to the four octocoral conserved gene blocks (block 1, red: *cox1–rns–nad1–cob*; block 2, blue: *nad6–nad3–nad4l*; block 3, yellow: *mtMutS–rnl–nad2–nad5–nad4*; block 4, green: *trnM–cox3–atp6–atp8–cox2*). Arrows indicate the direction of replication. Highlighted lines mark gene block transcript direction that differ from the ancestral gene order (A).

Table 5. Pennatulacea mitochondrial gene lengths (base pairs), intergenic regions (positive values) and overlapping genes (negative values) (base pairs) in specimens in this study with gene order B (*Anthoptilum* spp.), G (*Umbellula* sp. 1) and the bipartite genome (*Umbellula* sp. 3).

Genes	*Anthoptilum grandiflorum*	*Anthoptilum* sp. 1	Genes	*Umbellula* sp. 1	*Umbellula* sp. 3 (contig i)	*Umbellula* sp. 3 (contig ii)
cox2-cox1	**61**	**60**	***cox2-cox1***	**73**	**73**	
cox1	1566	1566	cox1	1569	1569	
cox1-rns	*37*	*37*	*cox1-rns*	*33*	*33*	
rns	1064	1059	rns	1060	1062	
rns-nad1	*53*	*52*	*rns-nad1*	*80*	*89*	
nad1	981	981	nad1	963	960	
nad1-cob	*−94*	*−94*	*nad1-cob*	*15*	*17*	
cob	1266	1266	cob	1215	1185	
cob-;nad6	*18*	*18*	*cob-nad6*	*−8*	*−8*	
nad6	555	555	nad6	579	579	
nad6-nad3	*63*	*64*	*nad6-nad2*	*13*	*13*	
nad3	354	354	nad2	1347	1347	
nad3-nad4L	*3*	*3*	*nad2-nad5*	*−13*	*−13*	
nad4L	294	294	nad5	1824	1824	
nad4L-nad4	*122*	*122*	*nad5-nad4*	*1*	*1*	
nad4	1449	1449	nad4	1449	1449	
nad4-nad5	*26*	*26*	*nad4-mtMutS /rnl-mtMutS*	*0*		*154**
nad5	1818	1818	mtMutS	3051		2940
nad5-nad2	*−13*	*−13*	*mtMutS-nad4L*	*6*		*6*
nad2	1350	1350	nad4L	294		294
nad2-rnl	*73*	*85*	*nad4L-nd3*	*3*		*3*
rnl	1799	1787	nad3	354		354
rnl-mtMutS	*154*	*154*	*nd3-rnl*	*64*		*114**
mtMutS	2946	2946	rnl	1763		1764
mtMutS-trnM	*187*	*186*	*rnl-trnM / nad4-trnM*	*293*	*74*	
trnM	71	71	trnM	71	71	
trnM-cox3	*55*	*58*	*trnM-cox3*	*44*	*44*	
cox3	786	786	cox3	660	660	
cox3-atp6	*57*	*57*	*cox3-atp6*	*178*	*178*	
atp6	708	708	atp6	708	708	
atp6-atp8	*24*	*24*	*atp6-atp8*	*17*	*17*	
atp8	216	216	atp8	216	216	
atp8-cox2	*42*	*42*	*atp8-cox2*	*15*	*15*	
cox2	762	762	cox2	777	777	
Total IGR (bp)	975	988		835	554	271
% IGR/mtgen	5.17	5.24		34.18	4.28	4.86
#ol genes	2	2		2	2	0
% pcg	80	80		80	87	53

Intergenic regions are represented in italic. Abbreviations: IGR, intergenic region; #ol genes, number of overlapping genes; pcg, protein-coding genes. OriH and OriL indicated by bold and underline respectively.
* possible origin of replication.

Table 6. Mitochondrial genome nucleotide composition of Pennatulacea of this study.

Species	A:	C:	G:	T:	GC:	GC Skew	AT Skew	%GC	Gene order
Anthoptilum grandiflorum	5974	3365	3579	5935	6944	0.03	0.00	36.8	B
Anthoptilum sp. 1	5977	3370	3582	5921	6952	0.03	0.00	36.9	B
Distichoptilum gracile	5624	3367	3869	6311	7236	0.07	−0.06	37.7	A
Funiculina quadrangularis	5630	3212	3720	6307	6934	0.07	−0.06	36.7	A
Halipteris cf. *finmarchica*	5534	3126	3633	6220	6759	0.08	−0.06	36.5	A
Kophobelemnon sp. 1	5617	3249	3730	6287	6979	0.07	−0.06	37.0	A
Kophobelemnon sp. 3	5688	3293	3772	6356	7065	0.07	−0.06	37.0	A
Kophobelemnon sp. 4	5720	3276	3774	6360	7050	0.07	−0.05	36.9	A
Pennatula aculeata	5523	3246	3724	6222	6970	0.07	−0.06	37.2	A
Pennatula cf. *aculeata*	5523	3251	3717	6224	6968	0.07	−0.06	37.2	A
Pennatula cf. *inflata*	5624	3338	3825	6340	7163	0.07	−0.06	37.4	A
Pennatula grandis	5593	3301	3786	6293	7087	0.07	−0.06	37.4	A
Protoptilum carpenteri	5538	3234	3712	6245	6946	0.07	−0.06	37.1	A
Umbellula huxleyi	5633	3232	3760	6302	6992	0.08	−0.06	36.9	A
Umbellula sp. 1	5479	3498	3572	6165	7070	0.01	−0.06	37.8	G (new)
Umbellula sp. 2	5592	3189	3705	6280	6894	0.07	−0.06	36.7	A
Umbellula sp. 3 _ 13k	3709	2295	2576	4358	4871	0.06	−0.08	37.6	Contig i
Umbellula sp. 3 _ 5k	1717	1004	1189	1669	2193	0.08	0.01	39.3	Contig ii
Virgularia mirabilis	5582	3195	3686	6307	6881	0.07	−0.06	36.7	A

Analyses of base composition

Overall base composition (G + C) ranged from 36.5% (*Halipteris* cf. *finmarchica*) to 37.8% (*Umbellula* sp. 1) (Table 6), similar to other octocorals (range = 36–38%). We also calculated GC-Skew and AT-Skew and found specimens with the same gene order to have similar strand-specific compositional asymmetry (Table 6). *Anthoptilum* species (gene order B) have the most even distribution with lower values of skew, similar to Isididae sp. BAL208-1 (Brugler and France 2008).

No tandem repeats were found in any genome. It seems possible that this result is due to the assembly cleaning

Table 7. Presence of inverted repeats and A-T rich stem-loop as evidence for OriL.

Gene order	Species	Inverted repeats	A-T rich stem loop
B	*Anthoptilum grandiflorum*		x
B	*Anthoptilum* sp. 1		x
A	*Distichoptilum gracile*		
A	*Funiculina quadrangularis*	x	x
A	*Halipteris* cf. *finmarchica*		
A	*Kophobelemnon* sp. 1		x
A	*Kophobelemnon* sp. 3		
A	*Kophobelemnon* sp. 4	x	x
A	*Pennatula aculeata*	x	x
A	*Pennatula* cf. *aculeata*	x	x
A	*Pennatula* cf. *inflata*	x	x
A	*Pennatula grandis*	x	x
A	*Protoptilum carpenteri*	x	x
A	*Umbellula huxleyi*		
new - G	*Umbellula* sp. 1		x
A	*Umbellula* sp. 2	x	
contig i	*Umbellula* sp. 3 _ 13k		
contig ii	*Umbellula* sp. 3 _ 5k		N/A
A	*Virgularia mirabilis*	x	x

In gene order A, these are in the *nad4-trnM* region, for gene order B in the *nad4l-nad4*, and for gene order G, the *nad4-mtMutS* region.

process. Inverted repeats were present in nine genomes, all of which had gene order A (Table 7). Inverted repeats were located between *nad4* and *trnM*, which is the largest IGR among our samples.

DNA walker (Figure 2) shows five abrupt changes in base composition bias in all samples in the three Pennatulacea gene arrangements. Three of these inversions are present in the middle of the *mtMutS, rns* and *rnl* genes, the other two are at intergenic regions. For gene order A, these occur in the intergenic regions *cox2-cox1* and *nad4-trnM*. For the two *Anthoptilum* species (gene order B), they are present in the IGRs *cox2-cox1* and *nad4L-nad4*. For gene orders A and B, the abrupt changes in base composition bias in these IGRs occur where a reversal in transcription orientation is present. The new gene arrangement (gene order G), found in *Umbellula* sp. 1, shows abrupt changes in base composition bias in the IGRs *cox2-cox1* and *rnl-trnM*. The intergenic region *rnl-trnM* (the largest in the *Umbellula* sp. 1 genome) does not coincide with a reversal in transcription orientation region. Instead, the reversal occurs between the junction of *rnl* (part of block 3) and gene block 4 (Figure 1). In *Umbellula* sp. 3, recovered as two contigs, abrupt inversions occur in its larger contig (13 kbp) in those same regions as found in gene order A (intergenic regions *cox2-cox1* and *nad4-trnM*). The shorter contig (5kbp) had an abrupt inversion before and after *rnl*.

MFold reveals stable stem-loop structures in all gene orders. A-T rich stem-loops were found in all samples at the *cox1-cox2* IGR, and in most of the other intergenic regions where abrupt changes in base composition bias were present (e.g. *nad4-trnM* in the gene order A; *nad4L-nad4* in gene order B; *rnl-trnM* in the new gene order ('G') (Table 7).

Given the gene order, conserved block boundaries, A-T rich regions, abrupt changes in base composition bias, size of intergenic region and stable A-T rich stem-loop structures described above for most of the species, we suggest that for all our samples and gene orders the oriH occurs in the intergenic region between *cox2-cox1*.

However, the oriL appears to differ by gene arrangement. For species with the same gene arrangement as the octocoral ancestral (gene order A), we suggest that the oriL is located in the intergenic region *nad4-trnM*. It represents one of the largest IGRs, and 64% of the samples with gene order A have inverted repeats in this region, as well as A-T rich stem-loop structures (Table 7). For the same reasons, apart from the presence of inverted repeats, the oriL of *Anthoptilum* spp. is most likely at the IGR *nad4L-nad4* as suggested for Isididae (Brugler and France 2008). For *Umbellula* sp. 1 (new gene arrangement, gene order G), the most likely oriL is the intergenic *rnl-trnM*, which is the junction of *rnl* (part of block 3) and gene block 4 and contains an abrupt change in base composition bias. The larger contig of *Umbellula* sp. 3 appears to have the same origin of replication pattern as found in gene order A (oriH: between *cox2-cox1;* oriL: between *nad4-trnM)*. In the smaller contig of *Umbellula* sp. 3, the most likely origin of replication is in the IGRs that have boundaries with *rnl* because of the presence of an abrupt change in base composition bias.

Protein-coding genes, ribosomal RNA genes and tRNAs

Gene lengths were similar among species (Tables 5 & 6). Gene size variations were found mainly in three species of *Umbellula* (*Umbellula* sp. 1, *Umbellula* sp. 3 and *Umbellula huxleyi*) in the *cob, cox2* and *cox3* loci. To better compare among studies, we applied our annotation method to all octocoral genomes available. *Umbellula huxleyi* has the longest version of *cob* (1593 bp), approximately 300 bp larger than the other Pennatulacea or octocoral species. This species also has one of the longest *cox3* genes (828 bp) among octocorals, although the size is not dissimilar to that seen in two species of Calcaxonia, *Junceella fragilis* (NC024181 - 837 bp) and bamboo coral BAL208-1 (NC010764 - 801 bp), when the same annotation method (editing of MITOS annotation to reflect ORFs) is applied. Protein-coding genes comprised 78.62% (*Distichoptilum gracile*) to 81.45% (*Halipteris* cf. *finmarchica*) of the mitochondrial genome (Table 4), which is within the previously reported octocoral range (75–88%).
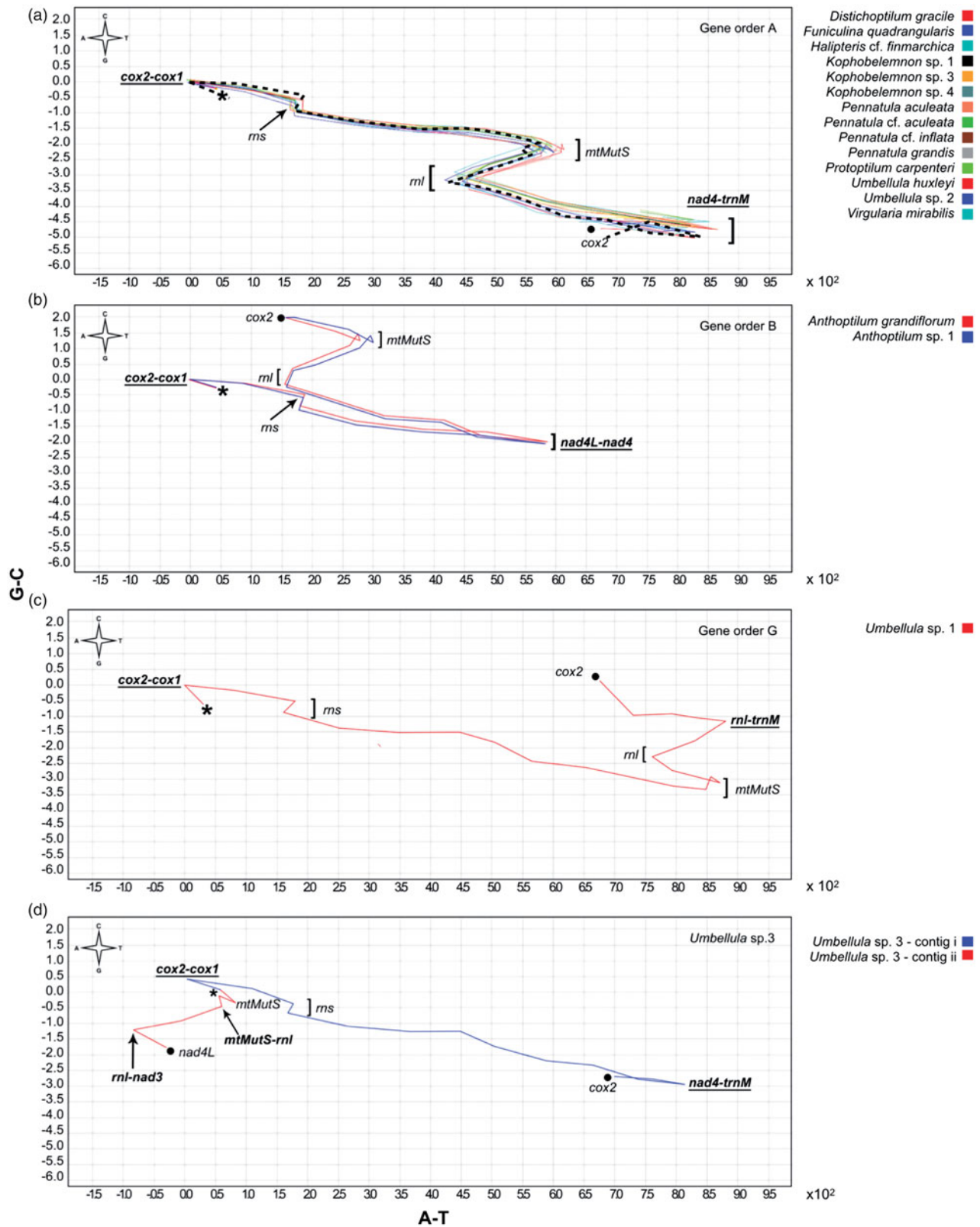
Figure 2. DNA-Walk (Graph DNA) vectorial representation of linearized pennatulacean mitochondrial genomes into a 2 D trajectory along the genome. Every nucleotide direction is determined by the compass in the left corner, representing a cumulative skew (window size: 900). **a:** gene order A. **b:** gene order B. **c:** new gene order G. **d:** bipartite genome. Genes and IGRs are marked in the abrupt inversions in base composition bias, which can be an indicative of origin of replication (suggested IGRs marked in **bold** and underline). Sequences start at the middle of *cox2* (*) and finish at the middle of *cox2* (.), except for *Umbellula* sp. 3 (contig ii) which starts at *mtMutS* (*) and finishes at *nad4L* (.). Dashed line represents the overall DNA-Walk shape for gene order A.
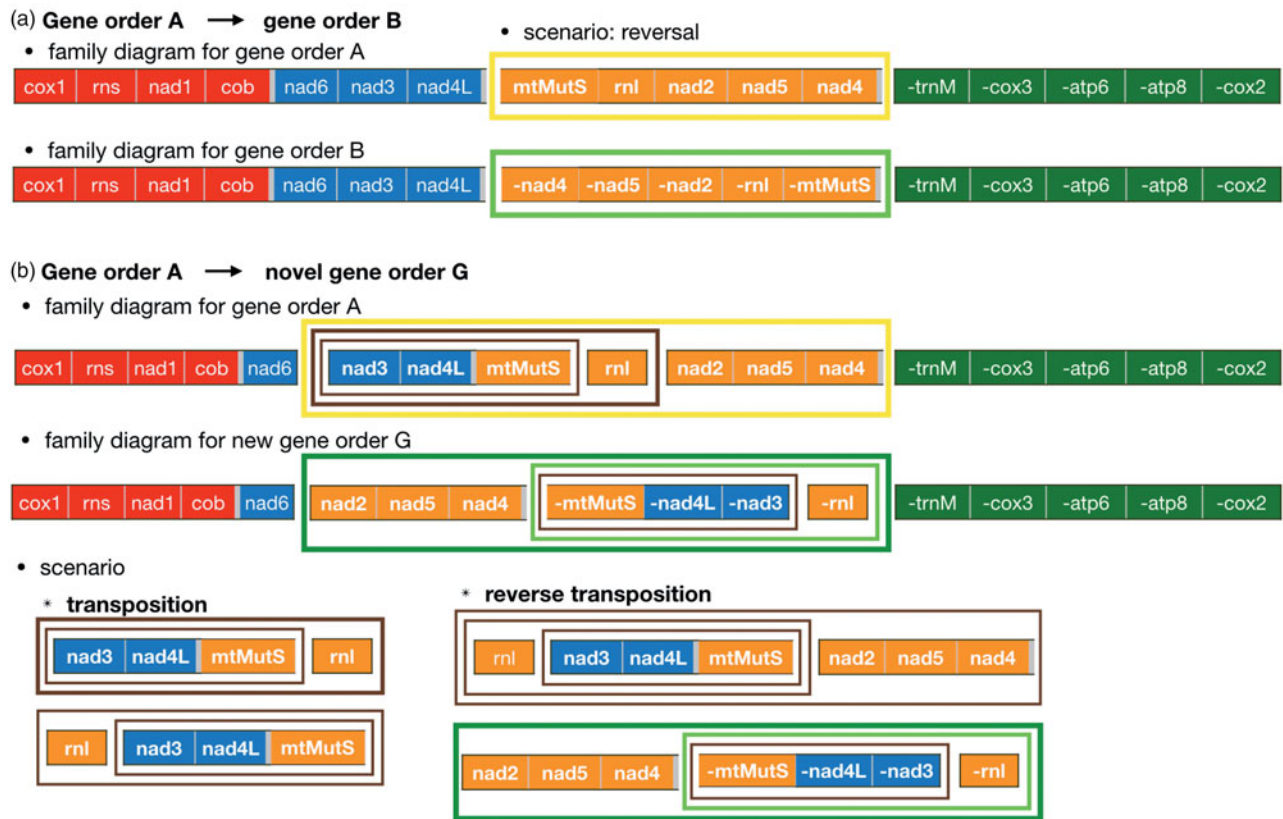
Figure 3. Scenarios of the possible mechanism of the mitochondrial gene rearrangement using parsimony hypothesis in CREx. **a.** Transition from gene order A to gene order B. **b.** Transition from gene order A to gene order G. Small coloured rectangles correspond to the four octocoral conserved gene blocks (block 1: red; block 2: blue; block 3: yellow; block 4: green). Highlighted coloured rectangles mark the genes that have been reversed (light green), transposed (brown), and reverse transposed (dark green).

Discussion

We sequenced the first deep-sea Pennatulacea mitochondrial genomes. Our data cover most of the morphotypes present in the Northeast Atlantic and provide new information about Pennatulacea, as well as Octocorallia and metazoans. The most important finds are: (1) Pennatulacea has at least three different gene arrangements; (2) gene order type B (Brugler and France 2008) is not exclusively present in bamboo corals (Octocorallia: Isididae); (3) the novel pennatulacean gene arrangement comprises a mix of genes from 'conserved gene blocks'; (4) one species has a bipartite genome, with strong evidence that both parts are circular.

Our results show that although 14 out of 18 of our samples have the same gene arrangement as the octocoral ancestral gene order, two genera, *Anthoptilum* and *Umbellula* (suborder Sessiliflorae), have different gene arrangements, indicating a higher diversity of gene arrangements in this group. Some studies have recorded greater diversity of gene arrangements and novel genetic structures in deep-sea taxa (Nakajima et al. 2016; Zhang et al. 2017, 2018), as well as a greater diversity of species in some deep-sea groups (Moura et al. 2012). This diversity may be related to the metabolic adaptations required to cope with an extreme environment (Zhang et al. 2018) and vertical divergences that could drive allopatric speciation in the abyssal depths (Taylor and Roterman 2017). Also, deep-sea pennatulaceans exhibit some morphological evolutionary changes relative to their shallow-water counterparts, which include a reduction in polyp number and an increase in polyp size (Williams 1995, 2011). *Anthoptilum* and *Umbellula* are two extremely deep dwellers in the Northeast Atlantic. *Umbellula* is the deepest known genus of sea pen, reaching depths of more than six thousand meters (*Umbellula thompsoni* 6240 m) (Molodtsova et al. 2008; Williams 2011). Our deepest sample was an *Umbellula* (*Umbellula* sp. 3) collected at 4173 m depth in the Whittard Canyon. *Anthoptilum* has an extensive bathymetric distribution (Williams 2011), represented here by our second deepest sample, collected at 3,148 m, which is representative of the deepest part of the known range of this genus (Williams 2011).

The two species of *Anthoptilum* have the same gene order as known from the octocoral family Isididae (gene order B) (Brugler and France 2008; van der Ham et al. 2009). Intramolecular recombination with inversion of the conserved block 2 has been suggested as the possible mechanism to generate this gene arrangement, assuming that it has evolved from the presumed ancestral gene order (Brugler and France 2008; van der Ham et al. 2009; Brockman and McFadden 2012). Our analyses in CREx (Figure 3a) support this scenario. Prior to this study, gene order B was considered specific to isidids (suborder Calcaxonia) (Brugler and France 2008), based on two mitochondrial genomes sequenced and extensive sequencing targeted at gene junctions of 100 octocorals covering a wide taxonomic spread (McFadden et al. 2006; Brugler and France 2008).

McFadden et al. (2006) attempted to sequence the *nad4l-mtMutS* gene boundary of 28 families and more than 100 specimens of octocorals, including 11 species of Pennatulacea with representatives from the deep-water genera *Umbellula* (*Umbellula* sp. NTM-C014384) and *Anthoptilum* (*Anthoptilum murrayi* - NTM-C014385) both from the Tasman Sea, Australia. Notably, attempts to amplify *Anthoptilum murrayi* were unsuccessful which, in light of our observations, might have resulted from a non-ancestral gene order. This fits the pattern found for *Paraminabea aldersladei*, where amplification failed for McFadden et al. (2006), and a divergent gene order (gene order E) was subsequently revealed (Brockman and McFadden 2012). Dolan et al. (2013) and Pante et al. (2012) have successfully amplified across the *nad4l-mtMutS* gene boundary in unidentified species of *Anthoptilum* (KF313832 and JN227921, respectively), although Dolan failed to obtain an amplification product in a second species. This suggests there may be more than one gene order in *Anthoptilum*.

Our results suggest that gene order B evolved at least twice in the Calcaxonia-Pennatulacea clade (McFadden et al. 2006). Similarly, gene order C, found in two morphospecies of *Anthomastus* as well as *Corallium* and *Hemicorallium* (Figueroa and Baco 2015), must have evolved more than once in the Scleraxonia-Alcyoniina clade. This supports the idea that, 'simple' gene rearrangements, may not be such rare and isolated evolutionary events (Boore 1999). They can occur and evolve in different branches of a clade, possibly through different evolutionary scenarios (e.g. shared characters, convergence), creating a higher genetic diversity. Furthermore, Figueroa and Baco (2015), when analyzing inverted repeats, showed that the apparently ancestral gene arrangement of the octocoral *Sibogagorgia cauliflora* is secondarily derived, amplifying our understanding of different scenarios and pathways that have led to the current octocoral gene arrangements.

Among all the gene orders observed within the octocorals, that of *Umbellula* sp. 1 is remarkably different (Figure 1). In five of the six previously described gene orders in octocorals, four large blocks of genes remain conserved (Brockman and McFadden, 2012) (conserved blocks illustrated in Figure 1), while the sixth appears not to have conserved the blocks (Pante et al. 2013). Our new gene order (G) provides a new scenario of gene arrangement, where two of these blocks are mixed. We tested different scenarios to illustrate how this could have been derived from all the known octocoral gene arrangements (data not shown). Under a parsimony hypothesis, gene order G is most likely derived from gene order A, since there are fewer steps in this transition and gene order A is the most common gene order found to date. We suggest gene order G arose through a transposition of *rnl*, followed by a reverse transposition of four genes (block 2: *nad3* and *nad4l*; and block 3: *mtMutS* and *rnl*) from two blocks previously described as conserved (Figure 3b).

This provides further evidence that the four blocks of genes previously considered conserved are not conserved in all species of octocorals, and suggests there are other mechanisms underlying gene arrangements through evolutionary time. Although block 2 appears to be conserved in Keratosa

(clade G1 of Demospongiae) and Medusozoa (Brockman and McFadden 2012), it is not in octocorals.

Another species of *Umbellula* (*Umbellula* sp. 3) also has a different pattern of arrangement, where we recovered two separate mitochondrial chromosomes. Multipartite mitochondrial genomes have been well documented in metazoans where the size and number of chromosomes vary remarkably (Watanabe et al. 1999; Wei et al. 2012; Lavrov and Pett 2016; Phillips et al. 2016; Kim et al. 2018). Despite the attempts to explain how and why a multipartite genome may have arisen, it is still unclear. Two main broad hypotheses have been discussed: genetic drift and/or some Darwinian evolutionary advantage (Rand 2009; Kim et al. 2018). Fragmented mitochondrial genomes usually result in small mitochondrial chromosomes. Natural selection could be one of the drivers for the predominance of small mitochondrial genomes in metazoans. For example, in some cnidarians and bilaterians, the mitochondrial genome has become more compact with a reduction in the number of genes, smaller intergenic regions and overlapped genes (Lang et al. 1999; Lavrov 2007; Shao et al. 2012), and in heteroplasmic crickets, smaller mitochondrial genomes are transmitted to offspring more frequently than larger ones (Rand and Harrison 1989). It has also been suggested that a multipartite genome might be more efficient in the process of gene expression (Rand 2009), which could potentially be an advantage in extreme environments such as the deep sea, where resources may sometimes be limited. In bacteria, multipartite genomes have been linked to niche-specialized functions, which have been associated with adaptions to new environments (diCenzo et al. 2018).

Cnidarians have been shown to have a high diversity of mitochondrial genome architecture, including circular, linear, and multiple linear genomes (Lavrov and Pett 2016). Yet, to date, multipartite circular genomes have been reported only in Bilateria (see review by Lavrov and Pett 2016). Even though sea pens are not included in the Bilateria, morphological bilateral symmetry is a common (but not uniform) characteristic of sea pen colonies (i.e. bilateral: *Pennatula*; radial: *Veretillum*). Morphological symmetry is also observed at the polyp level with bilateral symmetry often observed in the arrangement of the mesenteries in cnidarians (Berking 2007; Salvini-Plawen 2009). In fact, cnidarians not only have bilateral morphological traits but also some cnidarians have the *Hox* gene for the bilateral body plan (Finnerty et al. 2004; Dubuc et al. 2018). Therefore, they have also some of the genetic components that Bilateria have, so it is not so surprising that, perhaps, a cnidarian can have a multipartite circular genome.

Although hypotheses of possible mechanisms of gene arrangement in cnidarians and octocorals have been discussed in detail in some studies (Uda et al. 2011; Brockman and McFadden 2012; Kayal et al. 2012; Figueroa and Baco 2015; Lavrov and Pett 2016), only a small number of whole mitochondrial genomes are currently available, which limits interpretations. An important player in the mitochondrial genome architecture could be the *mtMutS* gene (Bilewitch and Degnan 2011; Brockman and McFadden 2012). Its role has been linked to DNA mismatch repair (Beagley et al. 1995; Bilewitch and Degnan 2011), including double-stranded break

repair, which could contribute to more conserved gene arrangements, but also could facilitate more variation through successful repairs.

Intergenic regions and origins of replications

In the three different pennatulacean gene arrangements, we hypothesize that the origins of replication in both strands occur in intergenic regions where abrupt changes in the direction of base composition bias occur.

In species with gene order A, B, G and in the large contig of our multipartite genome, our analyses suggest that oriH is between *cox2* and *cox1*, where there is a reversal in transcription orientation, as described for *Sarcophyton glaucum* (gene order A) and Isididae sp. BAL208-1 (gene order B) (Beaton et al. 1998; Brugler and France 2008).

Our analyses suggest that oriL occurs in large IGRs (Table 3), coinciding with the reversal in transcription orientation (in gene orders A and B), and/or at the intergenic regions downstream of *tRNA* (in gene order A and the novel gene order G). In some octocorals, oriL has been associated with the *trnM* intergenic region (Brugler and France 2008). However, the origins of replication have always occurred at the junctions between conserved gene blocks (Brockman and McFadden 2012), as we have found in the novel gene order G. Inverted repeats have proven to be important for the initiation of DNA replication in several organisms (Pearson et al. 1996). We found inverted repeats in most of the specimens with gene order A, and they were all located in the intergenic region between *nad4* and *trnM,* which is where we suggest the oriL is for this gene order (Table 2).

These results concur with our current understanding that in metazoans the origin of replication most likely proceeds from intergenic regions (Clayton 1991; Lewis et al. 1994; Brugler and France 2008). Origins of replication have been associated with different features, such as A + T rich sequences, stable stem-loops, and A-T rich stem-loops (Wolstenholme 1992; Pearson et al. 1996). We identified A-T rich stem-loops in all three gene arrangements. A-T rich regions can cause instabilities in DNA sequences, because they form weaker pair bonds compared to CG bonds (Mohajeri and Nobandegani 2008), and they are also associated with a high mutation rate (Hamilton et al. 2017).

The origin of replication of both strands in *Umbellula* sp. 3 contig i follows the same pattern that is present in gene order A. In contig ii, the origin of replication is most likely located in the intergenic regions that have borders with *rnl*. However, although there are secondary structures and a change in base composition bias at these intergenic regions (*mtMutS-rnl* and *rnl-nad3*), we could not determine the exact location of the origin of replication.

Overall, few studies have discussed the origin of replication in octocorals (Beaton et al. 1998; Brugler and France 2008; Uda et al. 2011; Brockman and McFadden 2012) and none has discussed it in Pennatulacea.

In conclusion, the data presented here demonstrate that the evolution of the mitochondrial genome in Pennatulacea has been more varied than previously thought, at least in deep-sea environments. When more mitochondrial genome data become available, we will probably have a better picture of the architecture, evolution and the mechanisms driving the mitochondrial gene arrangements in Pennatulacea, Octocorallia and in Metazoa. We intend to supplement the mitochondrial genomes described here with a substantial amount of nuclear data in a subsequent work to explore the evolutionary history of Pennatulacea through phylogenetic analyses.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Raissa I. Hogan 　 http://orcid.org/0000-0003-4103-3020
Andrew J. Wheeler 　 http://orcid.org/0000-0001-7107-2229
A. Louise Allcock 　 http://orcid.org/0000-0002-4806-0040
Chris Yesson 　 http://orcid.org/0000-0002-6731-4229

References

France S, Hoover L. 2002. DNA sequences of the mitochondrial COI gene have low levels of divergence among deep-sea octocorals (Cnidaria: Anthozoa). Hydrobiologia 471:149–155.

Andrews S. 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Baillon S, Hamel JF, Mercier A. 2014. Diversity, distribution and nature of faunal associations with deep-sea pennatulacean corals in the northwest Atlantic. PLoS One. 9:e111519.

Baillon S, Hamel JF, Wareham VE, Mercier A. 2012. Deep cold-water corals as nurseries for fish larvae. Front Ecol Environ. 10:351–356.

Beagley CT, Macfarlane JL, Pont-Kingdon GA, Okimotopp R, Okada DRW. 1995. Mitochondrial genomes of anthozoa (Cnidaria). Prog Cell Res. 5: 149–153.

Beaton MJ, Roger AJ, Cavalier-Smith T. 1998. Sequence analysis of the mitochondrial genome of Sarcophyton glaucum: Conserved gene order among octocorals. J Mol Evol. 47:697–708.

Benson G. 1999. Tandem repeats finder:a program to analyze DNA sequences. Nucleic Acids Res. 27:573–580.

Berking S. 2007. Generation of bilateral symmetry in Anthozoa: A model. J Theor Biol. 246:477–490.

Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler PF. 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol. 69:313–319.

Bernt M, Merkle D, Ramsch K, Fritzsch G, Perseke M, Bernhard D, Schlegel M, Stadler PF, Middendorf M. 2007. CREx: Inferring genomic rearrangements based on common intervals. Bioinformatics. 23: 2957–2958.

Berntson EA, Bayer FM, Mcarthur AG, France SC. 2001. Phylogenetic relationships within the Octocorallia (Cnidaria: Anthozoa) based on nuclear 18S rRNA sequences. Mar Biol. 138:235–246.

Bilewitch JP, Degnan SM. 2011. A unique horizontal gene transfer event has provided the octocoral mitochondrial genome with an active mismatch repair gene that has potential for an unusual self-contained function. BMC Evol Biol. 11: 228.

Boore J. 1999. Animal mitochondrial genomes. Nucleic Acids Res. 27: 1767–1780.

Brockman SA, McFadden CS. 2012. The mitochondrial genome of Paraminabea aldersladei (Cnidaria: Anthozoa: Octocorallia) supports intramolecular recombination as the primary mechanism of gene rearrangement in octocoral mitochondrial genomes. Genome. Biol Evol. 4:994–1006.

Brugler MR, France SC. 2008. The mitochondrial genome of a deep-sea bamboo coral (Cnidaria, Anthozoa, Octocorallia, Isididae): genome structure and putative origins of replication are not conserved among octocorals. J Mol Evol. 67:125.

Clayton DA. 1991. Replication and transcription of vertebrate mitochondrial DNA. Annu Rev Cell Biol. 7:453–478.

De Clippele LH, Buhl-Mortensen P, Buhl-Mortensen L. 2015. Fauna associated with cold water gorgonians and sea pens. Cont Shelf Res. 105: 67–78.

diCenzo GC, Mengoni A, Perrin E. 2018 Emergence and features of the multipartite genome structure of the family Burkholderiaceae revealed through comparative and evolutionary genomics. bioRxiv 382267. Doi: https://doi.org/10.1101/382267

Dierckxsens, N, Mardulyn, P, Smits, G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Research 45:e18.

Dolan E, Tyler PA, Yesson C, Rogers AD. 2013. Phylogeny and systematics of deep-sea sea pens (Anthozoa: Octocorallia: Pennatulacea). Mol Phylogenet Evol. 69:610–618.

Dubuc TQ, Stephenson TB, Rock AQ, Martindale MQ. 2018. Hox and Wnt pattern the primary body axis of an anthozoan cnidarian before gastrulation. Nat Commun. 9:1–12.

Figueroa D, Baco A. 2013. Complete mitochondrial genomes elucidate phylogenetic relationships of the deep-sea octocoral Families Coralliidae and Paragorgiidae. Deep Sea Res Part II Top Stud.99:1–9.

Figueroa DF, Baco AR. 2015. Octocoral mitochondrial genomes provide insights into the phylogenetic history of gene order rearrangements, order reversals, and cnidarian phylogenetics. Genome Biol Evol. 7: 391–409.

Finnerty JR, Pang K, Burton P, Paulson D, Martindale MQ. 2004. Origins of bilateral symmetry: Hox and dpp expression in a sea anemone. Science. 304:1335–1337.

France, SC, Hoover, LL. 2002. DNA sequences of the mitochondrial COI gene have low levels of divergence among deep-sea octocorals (Cnidaria: Anthozoa). Hydrobiologia 471:149–155.

Gordon, A. Hannon, G.2010. Fastx-toolkit. FASTQ/A short-reads pre-processing tools. http://hannonlab/cshl edu/fastx_toolkit.

Hahn C, Bachmann L, Chevreux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads - A baiting and iterative mapping approach. Nucleic Acids Res. 41:e129.

Hamilton WL, Claessens A, Otto TD, Kekre M, Fairhurst RM, Rayner JC, Kwiatkowski D. 2017. Extreme mutation bias and high AT content in Plasmodium falciparum. Nucleic Acids Res. 45:1889–1901.

Jühling F, Pütz J, Bernt M, Donath A, Middendorf M, Florentz C, Stadler PF. 2012. Improved systematic tRNA gene annotation allows new insights into the evolution of mitochondrial tRNA structures and into the mechanisms of mitochondrial genome rearrangements. Nucleic Acids Res. 40:2833–2845.

Kayal E, Bentlage B, Collins AG, Kayal M, Pirro S, Lavrov DV. 2012. Evolution of linear mitochondrial genomes in medusozoan cnidarians. Genome Biol Evol. 4:1.

Kayal E, Lavrov DV. 2008. The mitochondrial genome of Hydra oligactis (Cnidaria, Hydrozoa) sheds new light on animal mtDNA evolution and cnidarian phylogeny. Gene. 410:177.

Kayal E, Roure B, Philippe H, Collins AG, Lavrov DV. 2013. Cnidarian phylogenetic relationships as revealed by mitogenomics. BMC Evol Biol. 13:5.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28: 1647–1649.

Kerr, R. 1824. A general history and collection of voyages and travels arranged in systematic order, forming a complete history of the origin and progress of navigation, discovery, and commerce, by sea and land, from the earliest ages to the present time, Vol 8. Edinburgh: W. Blackwood. pp 338.

Kim T, Kern E, Park C, et al. 2018. The bipartite mitochondrial genome of Ruizia karukerae (Rhigonematomorpha, Nematoda). Sci Rep. 8:4–11.

Kushida Y, Reimer JD. 2018. Molecular phylogeny and diversity of sea pens (Cnidaria: Octocorallia: Pennatulacea) with a focus on shallow water species of the northwestern Pacific Ocean. Mol Phylogenet Evol. 131:1–12.

Lang BF, Gray MW, Burger G. 1999. Mitochondrial Genome Evolution and the Origin of Eukaryotes. Annu Rev Genet. 33:351–397.

Lavrov DV. 2007. Key transitions in animal evolution: A mitochondrial DNA perspective. Integr Comp Biol. 47:734–743.

Lavrov DV, Pett W. 2016. Animal mitochondrial DNA as we do not know it: Mt-Genome organization and evolution in nonbilaterian lineages. Genome Biol Evol. 8:2896–2913.

Lewis DL, Farr CL, Farquhar a. L, Kaguni LS. 1994. Sequence, organization, and evolution of the A + T region of Drosophila melanogaster mitochondrial DNA. Mol Biol Evol. 11:523–538.

Lobry JR. 1996. A simple vectorial representation of DNA sequences for the detection of replication origins in bacteria. Biochimie. 78:323–326.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 17:10.

McFadden CS, France SC, Sánchez JA, Alderslade P. 2006. A molecular phylogenetic analysis of the Octocorallia (Cnidaria: Anthozoa) based on mitochondrial protein-coding sequences. Mol Phylogenet Evol. 41: 513–527.

Mohajeri A, Nobandegani FF. 2008. Detection and evaluation of hydrogen bond strength in nucleic acid base pairs. J Phys Chem A. 112: 281–295.

Molodtsova TN, Sanamyan NP, Keller NB. 2008. Anthozoa from the northern Mid-Atlantic Ridge and Charlie-Gibbs Fracture Zone. Mar Biol Res. 4:112–130.

Moura CJ, Cunha MR, Porteiro FM, Yesson C, Rogers AD. 2012. Evolution of Nemertesia hydroids (Cnidaria: Hydrozoa, Plumulariidae) from the shallow and deep waters of the NE Atlantic and western Mediterranean. Zool Scr. 41:79–96.

Nakajima Y, Shinzato C, Khalturina M, Nakamura M, Watanabe H, Satoh N, Mitarai S. 2016. The mitochondrial genome sequence of a deep-sea, hydrothermal vent limpet, Lepetodrilus nux, presents a novel vetigastropod gene arrangement. Mar Genomics. 28:121–126.

Pante E, Saucier EH, France SC. 2013. Molecular and morphological data support reclassification of the octocoral genus Isidoides. Invertebr Syst. 27:365–378.

Pante E, France SC, Couloux A, Cruaud C, McFadden CS, Samadi S, Watling L. 2012. Deep-sea origin and in-situ diversification of chrysogorgiid octocorals. PLoS One. 7:e38357.

Park, E, Kim, B, Won, Y-J. 2010. The complete mitochondrialgenome of Dendronephthya gigantea (Anthozoa: Octocorallia: Nephtheidae). Korean J. Syst. Zool 26:197–201.

Park E, Hwang D-S, Lee J-S, Song J-I, Seo T-K, Won Y-J. 2012. Estimation of divergence times in cnidarian evolution based on mitochondrial protein-coding genes and the fossil record. Mol Phylogenet Evol. 62: 329–345.

Park E, Song JI, Won YJ. 2011. The complete mitochondrial genome of Calicogorgia granulosa (Anthozoa: Octocorallia): Potential gene novelty in unidentified ORFs formed by repeat expansion and segmental duplication. Gene. 486:81–87.

Pearson CE, Zorbas E, Price H, Gerald B, Zannis-Hadjopoulos M. 1996. Inverted repeats, stem-loops, and cruciforms: significance for initiation of DNA replication. J Cell Biochem. 63:1–22.

Pérez CD, Neves BDM, Cordeiro RT, Williams GC, Cairns SD. 2016. Diversity and Distribution of Octocorallia. In: Goffredo S, Dubinsky Z, editor. The Cnidaria, Past, Present and Future. The world of Medusa and her sisters. Springer: 109–123.

Perna NT, Kocher TD. 1995. Patterns of nucleotide composition at four-fold degenerate sites of animal mitochondrial genomes. J Mol Evol. 41:353–358.

Phillips WS, Brown AMV, Howe DK, Peetz AB, Blok VC, Denver DR, Zasada IA. 2016. The mitochondrial genome of Globodera ellingtonae is composed of two circles with segregated gene content and differential copy numbers. BMC Genomics. 17:1–14.

Pont-Kingdon G, Okada NA, Macfarlane JL, Beagley CT, Watkins-Sims CD, Cavalier-Smith T, Clark-Walker GD, Wolstenholme DR. 1998. Mitochondrial DNA of the coral Sarcophyton glaucum contains a gene for a homologue of bacterial MutS: a possible case of gene transfer from the nucleus to the mitochondrion. J Mol Evol. 46:419.

Rand DM. 2009. "Why genomes in pieces?" revisited: Sucking lice do their own thing in mtDNA circle game. Genome Res. 19:700–702.

Rand DM, Harrison RG. 1989. Molecular population genetics of mtDNA size variation in crickets. Genetics. 121:551–569.

Rice P, Longden L, Bleasby A. 2000. EMBOSS: The European molecular biology open software suite. Trends Genet. 16:276–277.

Salvini-Plawen LV. 2009. On the origin and evolution of the lower Metazoa. Zeitschrift Fuer Zool Syst Und Evol. 16:40–87.

Sánchez, JA, McFadden, CS, France, SC, Lasker, HR. 2003. Molecular phylogenetic analyses of shallow-water Caribbean octocorals. Marine Biology 142:975–987.

Shao R, Zhu XQ, Barker SC, Herd K. 2012. Evolution of extensively fragmented mitochondrial genomes in the lice of humans. Genome. Biol Evol. 4:1088–1101.

Taylor ML, Roterman CN. 2017. Invertebrate population genetics across Earth's largest habitat: The deep-sea floor. Mol Ecol. 26:4872.

Thomas JM, Horspool D, Brown G, Tcherepanov V, Upton C. 2007. GraphDNA: A Java program for graphical display of DNA composition analyses. BMC Bioinformatics. 8:21–28.

Uda K, Komeda Y, Fujita T, Iwasaki N, Bavestrello G, Giovine M, Cattaneo-Vietti R, Suzuki T. 2013. Complete mitochondrial genomes of the Japanese pink coral (Corallium elatius) and the Mediterranean red coral (Corallium rubrum): A reevaluation of the phylogeny of the family Coralliidae based on molecular data. Comp. Biochem Physiol - Part D Genomics Proteomics. 8:209–219.

Uda K, Komeda Y, Koyama H, Koga K, Fujita T, Iwasaki N, Suzuki T. 2011. Complete mitochondrial genomes of two Japanese precious corals, Paracorallium japonicum and Corallium konojoi (Cnidaria, Octocorallia, Coralliidae): Notable differences in gene arrangement. Gene. 476:27.

van der Ham JL, Brugler MR, France SC. 2009. Exploring the utility of an indel-rich, mitochondrial intergenic region as a molecular barcode for bamboo corals (Octocorallia: Isididae). Mar Genomics. 2:183–192.

Watanabe KI, Bessho Y, Kawasaki M, Hori H. 1999. Mitochondrial genes are found on minicircle DNA molecules in the mesozoan animal Dicyema. J Mol Biol. 286:645–650.

Wei DD, Shao R, Yuan ML, et al. 2012. The multipartite mitochondrial genome of liposcelis bostrychophila: Insights into the evolution of mitochondrial genomes in bilateral animals. PLoS One. 7:1–12.

Williams GC. 1995. The enigmatic sea pen genus Gyrophyllum — a phylogenetic reassessment and description of G. sibogae from Tasmanian waters (Coelenterata: Octocorallia). Proc Calif Acad Sci. 48: 315–328.

Williams GC. 2011. The global diversity of sea pens (Cnidaria: Octocorallia: Pennatulacea). PLoS One. 6:e22747.

Williams GC, Alderslade P. 2011. Three new species of pennatulacean octocorals with the ability to attach to rocky substrata (Cnidaria: Anthozoa: Pennatulacea). Zootaxa. 48:33–48.

Williams GC, Hoeksema BW, van Ofwegen LP. 2012. A fifth morphological polyp in pennatulacean octocorals, with a review of polyp polymorphism in the genera pennatula and pteroeides (anthozoa: Pennatulidae). Zool Stud. 51:1006–1017.

Wolstenholme DR. 1992. Animal Mitochondrial DNA: Structure and Evolution. Int Rev Cytol. 141:173–216.

Yesson C, Taylor ML, Tittensor DP, Davies AJ, Guinotte J, Baco A, Black J, Hall-Spencer JM, Rogers AD. 2012. Global habitat suitability of cold-water octocorals. J Biogeogr. 39:1278–1292.

Zhang Y, Sun J, Rouse GW, Wiklund H, Pleijel F, Watanabe HK, Chen C, Qian P-Y, Qiu J-W. 2018. Phylogeny, evolution and mitochondrial gene order rearrangement in scale worms (Aphroditiformia, Annelida). Mol Phylogenet Evol. 125:220–231.

Zhang B, Zhang Y-H, Wang X, Zhang H-X, Lin Q. 2017. The mitochondrial genome of a sea anemone Bolocera sp. exhibits novel genetic structures potentially involved in adaptation to the deep-sea environment. Ecol Evol. 7:4951–4962.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31:3406–3415.