Technical University of Denmark

DTU

Dispersal and speciation: The cross Atlantic relationship of two parasitic cnidarians

Dnyansagar, Rohit; Zimmermann, Bob; Moran, Yehu ; Praher, Daniela; Sundberg, Per; Møller, Lene Friis; Technau, Ulrich

Link back to DTU Orbit

DTU Library
Technical Information Center of Denmark

Dispersal and speciation: The cross Atlantic relationship of two parasitic cnidarians

Rohit Dnyansagar[a,1], Bob Zimmermann[a,1], Yehu Moran[a,b], Daniela Praher[a], Per Sundberg[c], Lene Friis Møller[d], Ulrich Technau[a,*]

[a] Department of Molecular Evolution and Development, University of Vienna, Austria
[b] Department of Ecology, Evolution and Behavior, Hebrew University of Jerusalem, Israel
[c] Department of Marine Sciences, University of Gothenburg, Sweden
[d] Danish Shellfish Centre, DTU Aqua, Technical University of Denmark, Denmark

ABSTRACT

How dispersal strategies impact the distribution of species and subsequent speciation events is a fundamental question in evolutionary biology. Sedentary benthic marine organisms, such as corals or sea anemones usually rely on motile larval stages for dispersal and therefore have a relatively restricted distribution along coasts. *Edwardsiella lineata* and *Edwardsiella carnea* are virtually indistinguishable edwardsiid sea anemones native to the east American and the Northern European coast, respectively. *E. lineata* is a facultative parasite to the ctenophore *Mnemiopsis leidyi*, while the life cycle of *E. carnea* is unknown. Recently *M. leidyi* was found in the Skagerrak carrying *Edwardsiella* sp. parasites, which raised the intriguing possibility that the invasive comb jellies acted as cargo for the facultative *E. lineata* parasites to establish a new population in Northern Europe. Here, we assessed the genetic differences between these two cryptic *Edwardsiella* species and isolated parasites from the invasive comb jelly *M. leidyi* in Sweden by comparing rRNA, whole transcriptomes, SNPs, ITS2 sequences and the gene complements of key developmental regulators, the Wnt gene family. We show that *E. carnea* and the parasite transcriptomes are more than 99% identical, hence demonstrating that *E. carnea* has a previously unknown parasitic life stage. ITS2 sequence analysis of *E. carnea* and *E. lineata* suggest that they may not be reproductively isolated. The transcriptomes of *E. lineata* and *E. carnea* are ~97% identical. We also estimate that the species diverged between 18.7 and 21.6 million years ago.

1. Introduction

How new species arise and how populations can spread over large distances and conquer new habitats is one of the fundamental questions in evolutionary biology and ecology. For marine benthic or sedentary organisms, the oceans can be unsurpassable hurdles, if the pelagic stage is either too short or not motile enough to cross the ocean. Among those animals are the Actiniaria (sea anemones), member of the class Anthozoa within the phylum Cnidaria. Cnidaria is the sister lineage to Bilateria consisting of very diverse body plans including sea anemones, corals and jellyfish. The Anthozoa, although commonly considered monophyletic, may actually represent a paraphyletic group, with the Hexacorallia (including Actiniaria) as the sister group to the Octocorallia plus Medusozoa (Kayal et al., 2013). The Anthozoa,

although commonly considered monophyletic, may actually represent a paraphyletic group, with the Hexacorallia (including Actiniaria) as the sister group to the Octocorallia plus Medusozoa (Kayal et al., 2013). Cnidaria is the sister lineage to Bilateria consisting of very diverse body plans including sea anemones, corals and jellyfish. While Medusozoa generally form both polyps and medusae, Anthozoa is characterized by the absence of the medusa stage, hence lacking a long-lived pelagic stage, which would allow the distribution over large distances (Bridge et al., 1992, 1995). The only pelagic stage of the sedentary sea anemones is the planula larva, which, depending on the species, lasts for a few days to several weeks before transforming into a polyp (Nyholm, 1943). The sea anemone *Nematostella vectensis* recently became one of the major model organisms among cnidarians for the study of comparative genomics, developmental biology and ecology (for reviews see
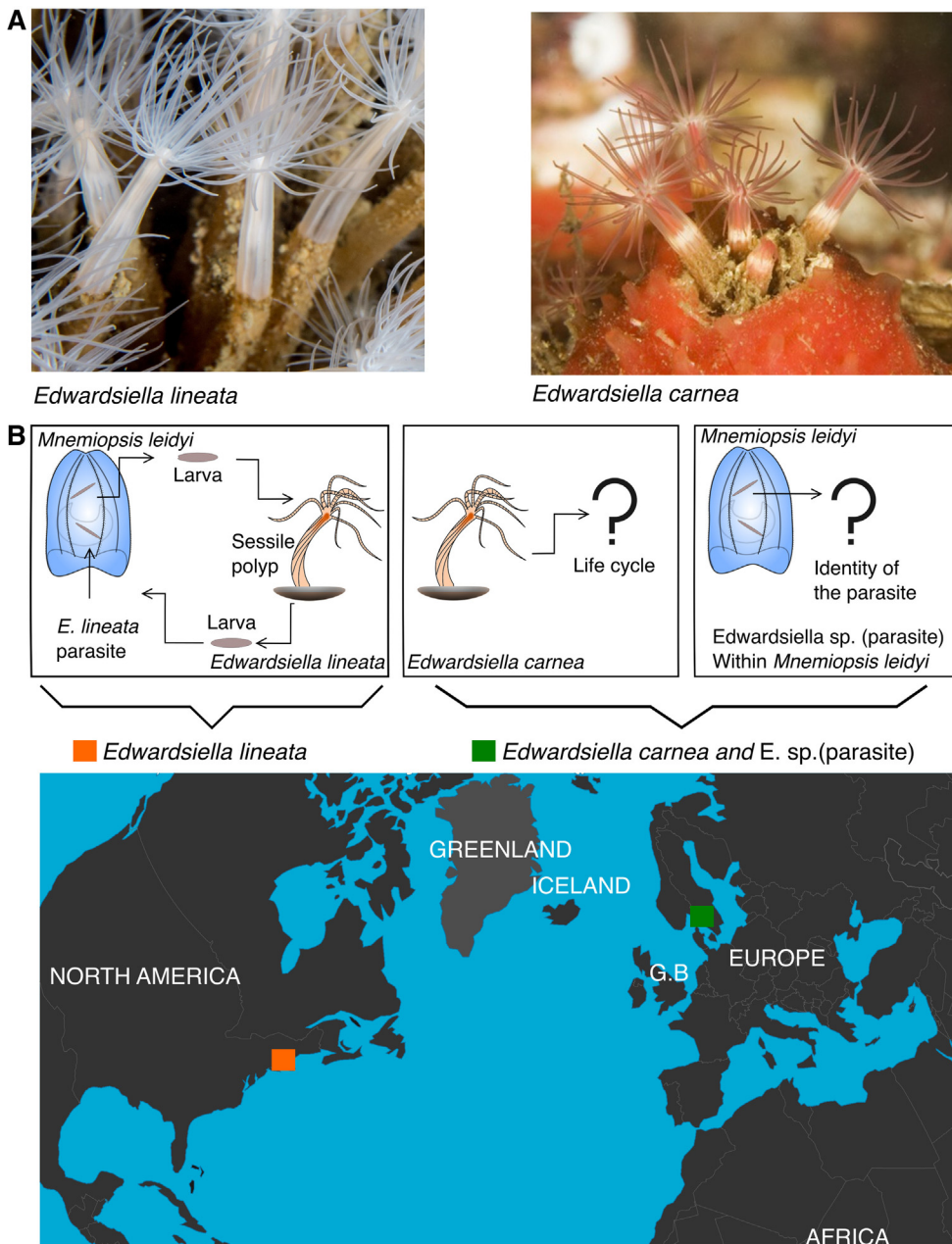
Fig. 1. A: *Edwardsiella lineata* (*Image credit: Alex Shure) Edwardsiella carnea* (Image credit: Kåre Telnes) and are two morphologically very similar sea anemones found at the opposite end of the Atlantic Ocean. *E. carnea* is found on Swedish West coast while *E. lineata* is found on American East coast. **B:** *E. lineata* is known to have a parasitic life cycle within the ctenophore *M. leidyi* as vermiform larvae. Outside of the ctenophore host *E. lineata* is able to develop into adult sessile polyp. Although *E. carnea* was described in 1856, the life cycle of the *E. carnea* is not fully understood. We also found *M. leidyi* ctenophore close to the Swedish west coast infected with the *Edwardsiidae* parasite.

Darling et al., 2005; Genikhovich and Technau, 2009; Layden et al., 2016; Rentzsch and Technau, 2016; Steele et al., 2011; Technau and Schwaiger, 2015) *N. vectensis* originates from the estuarine habitats along the east coast of America (Hand, 1990; Hand and Uhlinger, 1992; Darling et al., 2005) and it belongs to the family of *Edwardsiidae*. Sea anemones of this family have long slender bodies, which are buried in sediments or crevices in the rocks.

Two sea anemone species of this family, closely related to *Nematostella*, are *Edwardsiella lineata* and *Edwardsiella carnea* (Fig. 1). *E. lineata* is found at the east coast of North America. Interestingly, planula larvae of *E. lineata* can enter the gut of ctenophores, mostly *Mnemiopsis leidyi*, and transform into a worm-like parasitic stage without tentacles or mesenteries (Fig. 1B) (Crowell, 1976). Due to its facultative parasitic life cycle in *M. leidyi*, *E. lineata* has been studied in some detail (Reitzel et al., 2006; Reitzel et al., 2007; Reitzel et al., 2009) and its transcriptome is available (Stefanik et al., 2014). *E. carnea*, which is found along the Atlantic coast of Sweden and Norway (Gosse, 1856; Daly, 2002) is extremely similar morphologically to *E. lineata*. Indeed, they

differ only by two out of sixty morphological characters (Daly, 2002), raising the question of their phylogenetic relationship. *E. carnea* was first identified as *Edwardsia carnea* (Gosse, 1856). Since then the species has been renamed several times as *Halcampa microps* (Gosse, 1856), *Milneedwardsia carnea* (Carlgren, 1921) and *Fagesia carnea* (Delphy, 1938). The name *Edwardsiella carnea* is currently accepted by *The World Register for Marine Species,* which is based on *The Marine Fauna of the British Isles and North-West Europe.* Although *E. carnea* has been identified in 1856 (Daly et al., 2002; Gosse, 1856; Daly, 2002; Selander et al., 2009), virtually no information is available about its life cycle. In contrast to *E. lineata*, parasitic stages were never described for *E. carnea*.

The American ctenophore *M. leidyi* (Agassiz, 1865) was for the first time recorded in Northern Europe near Kiel in 2006 (Javidpour et al., 2006) and shortly thereafter in the North Sea (Faasse and Bayha, 2006; Boersma et al., 2007) and in the Skagerrak region of the Swedish West Coast (Hansson, 2006). Since then it has been found most years in Swedish waters (Selander et al., 2009). During the peak abundance of *M. leidyi* in 2007 endoparasitic sea anemone larvae were observed for

the first time attached inside the invasive ctenophore in Swedish waters (Selander et al., 2009) and have been frequently observed since then. In their native environment infection of the comb jellies by *E. lineata* is common (Crowell, 1976; Bumann and Puls, 1996), yet, the identity of the parasites found in the invasive ctenophore in Sweden remained unclear. It could be the parasitic stage of *E. lineata*, hosted and carried along by the invasive ctenophore, or, alternatively, represent a parasite belonging to *E. carnea* or another edwardsiid species. However, the comb jelly *Bolinopsis infundibulum* is also native to the Skagerrak area (Selander et al., 2009) and a single record in the literature also suggests the possibility that *E. carnea* may be a parasite of *B. infundibulum* (Stephenson, 1935). Identification of the parasite based on morphology is even more challenging as it is vermiform and devoid of most morphological features used to characterize sea anemones such as tentacles and mesenteries. Therefore, *E. lineata*, *E. carnea* and *E.* sp. (parasite) may form a species complex, consisting of closely related, morphologically almost indistinguishable species. These species complexes have the ability to showcase the successfully maintained morphology despite underlying variable genetics, behavior, or physiology. These species also reveal the limitations of morphology based traditional methods and accentuate the need of novel methods to classify organisms.

Commonly used methods to distinguish species include phylogenetic comparison of rRNA sequences, nuclear and mitochondrial genes, microsatellite genotyping, comparative genomics, and Internal Transcribed Spacer 2 compensatory base change analysis. In order to resolve the *Edwardsiella* species complex, we generated transcriptomes of *E. carnea* and the parasites isolated in Sweden and compared them to the published transcriptome dataset of *E. lineata* (Stefanik et al., 2014). Using a variety of methods our results show that *E. carnea* and *E. lineata* are genetically very similar, suggesting a very close species relationship. Moreover, we show that the parasite found in an American ctenophore at the shore of Sweden stems from *E. carnea* polyps, demonstrating that this species also has a parasitic life stage, which can enter foreign hosts.

2. Materials and methods

2.1. Collection of animals

E. carnea were collected by dredging at 40 m depth on the Swedish West Coast (58°°21′N, 11°06′E; 58°21′N, 11°07′E) August 2012. *E. carnea* were found at the bases of the soft coral *Alcyonium digitatum* and on rocks. The sea anemones were kept in sea water at Sven Loven Marine Center, Kristineberg, Fiskebäckskil and fed with cultured *Artemia* nauplii. They were starved for three days and then flash frozen in liquid nitrogen and kept in −80 °C until used. The ctenophore *M. leidyi* (Agassiz), infected by endoparasitic larvae, was collected in the Gullmar Fjord off the Swedish West Coast (58°21′N, 11°24′E). Larvae attached close to the pharynx or stomach of *M. leidyi* were gently removed from the ctenophore and preserved in RNAlater (Ambion).

2.2. Generation and quality control of transcriptome datasets

The *E. lineata* transcriptome and read data were obtained from EdwardsiellaBase (Stefanik et al., 2014). The read data consist of ~340 million paired end Illumina reads. Total RNA was extracted from four *E. carnea* adults using the VWR Omega-BioTek E.Z.N.A molluscan RNA isolation kit (catalog number R6875-00). Transcriptome sequencing of *E. carnea* was performed at GENEWIZ South Plainfield, NJ with Illumina Hiseq 2000, which resulted in ~180 million paired end reads of length 125. Total RNA was extracted from six *E.* sp. specimens with Trizol reagent according to the manufacturer's instructions. Library preparation and sequencing of the mRNA was performed at VBCF NGS unit (www.vbcf.ac.at) using an Illumina HiSeq 2500 sequencing system which resulted in ~660 million paired end reads. The quality control tool FASTQC for the high throughput sequencing data was used to assess the read quality. Due to the absence of genome data, de novo transcriptome assembly was performed with Trinity (Grabherr et al., 2011) in strand specific mode with the additional option of adapter filtering using Trimmomatic (Bolger et al., 2014). We used Transdecoder (Haas and Papanicolaou, 2012) utility to predict open reading frames (ORF) in transcripts and discarded transcripts of ORFs, which were less than 100 amino acids. To assess biological completeness of the transcriptomes, we subjected the transcriptomes to the BUSCO (Simão et al., 2015) analysis using 843 well-curated protein-coding genes. While the transcriptomes of *E.* sp. (parasite) and *E. lineata* transcriptomes contained 52% and 47% of the BUSCO marker genes, respectively, the *E. carnea* transcriptome is the most complete dataset among the transcriptomes under study with 92% (single copy and duplicated) marker genes (Supplementary Fig. 1). To construct the 'pseudo-genome' dataset, orthologous sequences were identified with OrthoMCL (Li et al., 2003) and used for further comparative genomic and variant analysis. Marker sequences such as 18S rRNA and *Wnt* genes were identified with blast homology search. Assembled transcriptomes and raw sequencing reads for *E. carnea* and *E.* sp. Parasite have been deposited to NCBI Transcriptome Shotgun Assembly (TSA) database under the accessions GGGD00000000 and GGGB00000000 respectively.

2.3. Analysis of sequence identity and genetic distance

We used the OrthoMCL pipeline (Li et al., 2003) to find a reliable set of orthologous sequences from *E. lineata*, *E. carnea*, *N. vectensis* and *E.* sp. (parasite). OrthoMCL uses homology searches with BLAST along with the Markov Cluster aLgorithm (MCL) to determine homologs. The resulting groups of homologs were further processed within house python script (https://github.com/dnyansagar/edwardsiella). In order to avoid comparisons between fragmented sequences, orthologous groups sharing less than 90% of the sequence length among all group members were filtered out. Pairwise alignments of the transcript sequences were done using lastz (Harris, 2007) and percent identity scores obtained for each alignment were used to calculate the average percentage identity between the datasets. To account for the size differences of transcripts being aligned, the '–noytrim' flag available in lastz algorithm was used, which extends alignment to the end of longer sequences. We also used the orthologous sequences obtained through OrthoMCL to calculate genetic distances using FDNADIST algorithm from the EMBOSS package (Rice et al., 2000). In order to assess the stability of the percent identity distance differences, 95% confidence intervals were calculated via bootstrapping the source alignments. Standard deviations from the mean were estimated from the distribution percent identities of 1000 bootstrap replicates of the alignments (Supplementary data 2). A calculation of the confidence intervals of the LogDet distances were estimated based on the approximate, normal theory confidence intervals as described in (Cai et al., 2015) as implemented in the heplots R package.

2.4. Protocol validation/verification

To verify our comparative genomics analysis protocol, we tested the protocol on well-studied group of apes. We obtained 14,068 orthologous groups via OrthoMCL pipeline from *Pan troglodytes, Homo sapiens, Gorilla gorilla* data from the Ensembl BioMart (Kinsella et al., 2011). Additionally, we created 12 pseudo transcriptomes from each of our read dataset, namely *E. lineata*, *E. carnea*, E. sp. (parasite) and then followed the same protocol of comparative genomics.

2.5. Phylogenetic analysis

For the phylogenetic comparison between the closest sea anemone species, we obtained 18S rRNA sequences of *Edwardsianthus gilbertensis*, *Edwardsia andresi*, *Edwardsia japonica*, *Edwardsia elegans*, *Edwardsia sipunculoides*, *Edwardsia timida*, *Edwardsia tuberculata* and *Nematostella*

vectensis from NCBI (See Supplementary data 3 for accession). 18S rRNA sequences of these species were aligned with MAFFT (Katoh and Standley, 2013) using the E-INS-i algorithm. The resulting alignment was subjected to the maximum likelihood phylogenetic analysis using IQ-TREE (Nguyen et al., 2015) with 1000 bootstrap samples. IQ-TREE tree topology and branch lengths were inferred under the Tamura-Nei (TN) substitution model with allowance for the proportion of the invariable sites (+I), which was selected with standard model selection (not including FreeRate models). The same alignment was used for the Bayesian phylogenetic analyses. Bayesian analysis with Mrbayes (Ronquist and Huelsenbeck, 2003) was run with the default nucleotide substitution model (4 × 4). Each analysis was set to run for 2×10^6 generations and every hundredth tree was sampled. MCMC generation was terminated when the standard deviation of split frequencies fell below 0.01. The first quartile of the sampled trees was discarded to assure better sample quality. The resulting tree has an identical topology to the tree obtained with maximum likelihood therefore the posterior support values for the branches are merged.

To search for *Wnt* sequences in the *E. lineata*, *E. carnea* and *E. sp.* (parasite) transcriptomes, *N. vectensis Wnt* sequences (Kusserow et al., 2005) were used as baits. The tblastx algorithm of NCBI Blast (Altschul et al., 1990) was used with 1e-3 as e-value cut-off. The sequences then subjected to MAFFT alignment with using the L-INS-i algorithm. The resulting alignment was edited manually using Jalview (Waterhouse et al., 2009). The resulting alignment was subjected to the maximum likelihood phylogenetic analysis using IQ-TREE (Nguyen et al., 2015). Best-fit model LG + I + G4 was chosen by the IQ-TREE according to BIC (Bayesian Information Criterion). The tree topology was confirmed with 500 bootstraps. The same alignment was also subjected to Bayesian phylogenetic analysis with MrBayes and the posterior probability mapped on the maximum likelihood tree.

2.6. Variant calling

For variant calling aligned reads were mapped to the orthologous gene set created earlier using bowtie2 with default settings (Langmead and Salzberg, 2012). The mapped reads were then subjected to the SAMtools mpileup program (Li et al., 2009), which provides a summary of the coverage of mapped reads on a reference dataset at a single base pair resolution. This summary was piped to VarScan (Koboldt et al., 2012) to call SNPs. A potential source of bias in the SNP analysis is the use of 'pseudo-genome' we created as a reference for the variant calling, however, we ensured with our strict filtering techniques for read data that only high quality reads will be used for variant calling and each SNP will have sufficient read support. Criteria used for filtering SNPs in order to avoid false positives are (i) read support for the SNP position should be more than 100 (ii) *p*-value for the SNP should be less than 0.01 (iii) Phred Quality Score for the base call should be more than 15 (call is > 90% accurate) Further, we used an in-house python script to compare the locations of SNPs in each transcript. We counted the occurrences of such events where transcripts from two species have SNP in same location.

2.7. Internal transcribed spacer 2 (ITS2)

Curated metazoan ITS2 sequences were downloaded from the ITS2 Database (http://its2.bioapps.biozentrum.uni-wuerzburg.de/) (Merget et al., 2012). Reads of *E. carnea*, *E. lineata* and *E. sp.* (parasite) were mapped against these sequences using bowtie2 (Langmead and Salzberg, 2012). RNA sequence for *N. vectensis* was obtained from RNAcentral database (Bateman et al., 2011). All the putative ITS2 sequences were subjected to the "Annotate" tool available at the ITS2 database to determine the boundaries of ITS sequences. ITS2 sequences were then subjected to RNAfold program (Lorenz et al., 2011) to predict the secondary structure of the ITS2. An option to predict best secondary structure based on minimum free energy and partition function was

selected from the RNAfold program. To compare and find compensatory base change (CBC), Input data in the XFasta format (fasta sequence with secondary structure) was prepared and used in 4SALE program (Seibel et al., 2008). To evaluate the efficiency of the method we applied the method to some of the known species groups such as Apes, Rodents and *Drosophila* (Supplementary data 1).

2.8. Divergence time estimates

For divergence time estimation, we followed the approach used by Peterson and colleagues (Peterson et al., 2004). In this approach seven conserved nuclear genes (ATP Synthase, Eukaryotic Translation Elongation Factor 1 Alpha 1, Methionine Adenosyltransferase 1A, Triosephosphate Isomerase, Catalase, Aldolase Fructose biphosphate, Phosphofructokinase) identified from taxa were used to calculate the divergence time of *Edwardsiella* species, with the poriferan *Oscarella carmela* as an outgroup. Sequences from *N. vectensis*, *E. carnea, E. lineata, E. sp.* parasite, *A. millipora* and *A. digitifera* were found via local BLASTP searches of the 15-taxon data set downloaded from GenBank accessions AY580167-AY580307 (Altschul et al., 1990). The transcriptome of *Oscarella carmela* was downloaded from http://www.compagen.org/ (Ereskovsky et al., 2017). The complements of *Drosophila melanogaster* and *Anopheles gambiae* were found via online NCBI BLAST searches ("Database Resources of the National Center for Biotechnology Information," 2017). Multi-way sequences from *N. vectensis*, *E. carnea, E. lineata, E. sp.* parasite, *A. millipora* and *A. digitifera* were found via local BLASTP searches (Altschul et al., 1990). lignments of individual genes were performed with MAFFT in E-INS-i mode (Katoh and Standley, 2013) and trimmed using trimAl in automated1 mode (Capella-Gutiérrez et al., 2009). A maximum likelihood tree was inferred using IQ-TREE using the model LG + I + G4 as selected using ModelFinder (Nguyen et al., 2015). Date estimates were determined using r8s version 1.81 using the Langley-Fitch likelihood method (Sanderson, 2003). Ranges were estimated by fixing the age of the bilaterian split between 555.0 and 641.7 Mya (dos Reis et al., 2015).

3. Results

3.1. 18s rRNA phylogenetic analysis shows close relationship of E. carnea, E. sp. (parasite) and E. lineata

In an effort to find species closely related to the model organism *Nematostella vectensis*, we collected four specimens of the edwardsiid sea anemone *E. carnea* on the coast of Lysekill, Sweden. Likewise, we pooled twelve specimens of parasites from the ctenophore *M. leidyi* collected off the coast of Sweden (Gullmar Fjord) and sequenced the transcriptomes of both. To assess the phylogenetic relationships of the two species and the parasite we carried out a phylogenetic analysis based on 18S rRNA sequences using both the maximum likelihood (IQ-TREE multicore version 1.5.5) (Nguyen et al., 2015) and Bayesian (MrBayes v3.2.6) (Ronquist and Huelsenbeck, 2003) approaches (Fig. 2). *Edwardsiella lineata, Edwardsiella carnea* and the *Edwardsiella* sp. parasite form a monophyletic clade with *N. vectensis*, within the family Edwardsiidae, supporting their close relationship. Additionally, upon examination of the alignments we found that *E. carnea* and *E. sp.* (parasite) are completely identical to each other, while *E. carnea* and *E. lineata* differ only by two base change across 1896 bases.

3.2. Assessment of genetic distance using comparison of 'pseudo-genomes'

Since rRNA is a strongly conserved molecule and may not reveal hidden variation underlying recent speciation events, we sought to analyze the whole transcriptome in more detail. Transcriptome data for *E. lineata* was recovered from published databases (Stefanik et al., 2014). Assemblies of *E. lineata*, *E. carnea* and *E. sp.* (parasite) comprise 117,890, 296,463 and 186,572 transcripts respectively (Supplementary
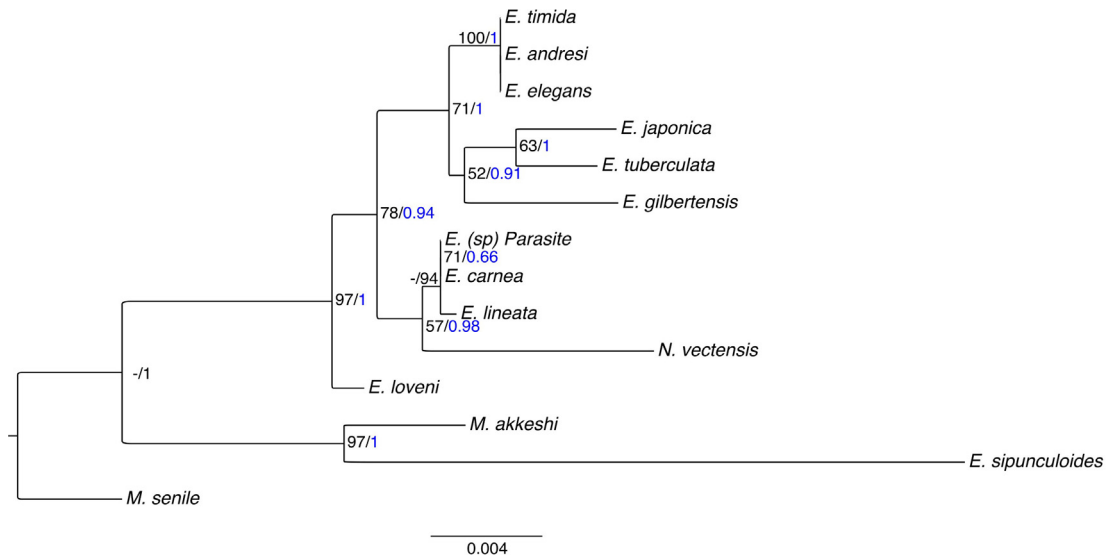
Fig. 2. Phylogenetic relationship among *Edwardsiidae* species based on 18S rRNA sequences. The phylogenetic tree was constructed with the maximum likelihood method implemented in IQ-TREE using *Metridium senile* as outgroup. The first value at the nodes indicates the Bootstrap support from the maximum likelihood method while the second value (shown in blue) indicates posterior probability support from the Bayesian inference method. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 2). We checked for the presence of BUSCO metazoan marker genes and found that the transcriptomes of *E.* sp. (parasite) and *E. lineata* transcriptomes contained between 81% (*E. lineata*) and 97% (*E. carnea*) of the gene set in partial or full-length form (Supplementary Fig. 1). This shows that all three transcriptomes are close to saturation.

In order to evaluate nucleotide-level sequence identity, de novo transcriptomes generally are not sufficient because of the potential mis-assembly of transcripts. Therefore, we constructed a 'pseudo-genome' from well-assembled transcripts, which were highly represented in the transcriptomes. We included *N. vectensis* as an outgroup reference to aid in comparative analysis of data. We selected 7021 orthologous sequence groups using the OrthoMCL pipeline (Li et al., 2003) and in-house Python scripts and created separate datasets for each species to pairwise compare the percentage identity and genetic distance with LogDet distance measure. Amongst several genetic distance measures available we selected LogDet distance due to its robustness to the composition biases (Massingham and Goldman, 2007) that may occur in transcriptome data (Zheng et al., 2011). The percentage identity measure calculated with lastz (Fig. 3) indicates that the transcriptomes of *E. carnea* and *E. lineata* are ~97% identical, while *E. carnea* and *E.* sp. (parasite) share more than 99% nucleotide identity. *E. carnea* and *E.* sp. (parasite) are also equidistant from *E. lineata* (~97%) and *N. vectensis* (~73%). Both *E. lineata* and *E. carnea* are ~73% identical to *N. vectensis*. The LogDet distance calculated with FDNADIST program from the EMBOSS (Rice et al., 2000) package is also shown in the Fig. 3. The FDNADIST program reads in DNA sequences and outputs a distance matrix. The LogDet distance calculated here between *E. carnea* and *E.* sp. (parasite) is 0.008, thus two orders of magnitude smaller than the distance between *E. carnea* and *E. lineata* (0.252). *N. vectensis* is the farthest from other species according to the measures in the study, although it is closer to *E. lineata* (0.793) than to *E. carnea* (1.126) or the *E.* sp. (parasite) (1.12612). This distinction in the relationship is not evident in the percentage sequence identity measure and thus provides a more effective measure to assess genetic differences among species. Our validation by dividing our datasets in twelve subsets shows very small standard deviation indicating the robustness of our comparative transcriptomic analysis protocol (Supplementary Fig. 3). To further validate our analysis, we also compared humans with chimpanzee and gorilla. The comparative genomic analysis between human and chimpanzee shows 99.03% identity and 0.013 LogDet distance. Human and gorilla
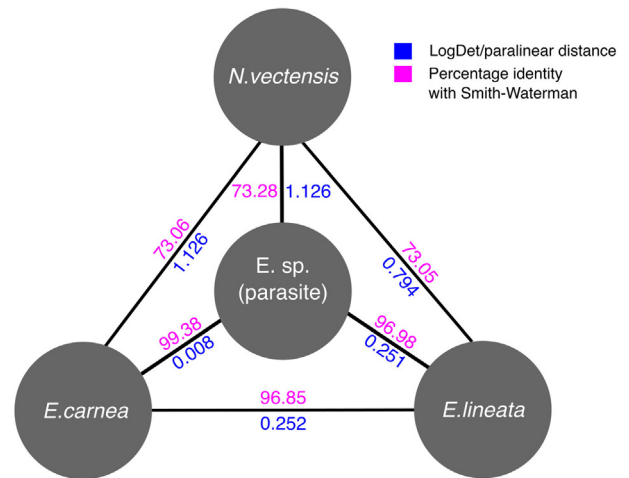

Fig. 3. A: Comparative transcriptomic analysis of *E. lineata*, *E. carnea*, *N. vectensis* and E. sp. (parasite). The LogDet distance (blue) measure of evolutionary divergence is calculated using the FDNADIST program from the EMBOSS package. The alignment of orthologous sequences used to create 'pseudo-genome' is used as input to FDNADIST. Percentage similarity (pink) calculated using Lastz program with same orthologous sequences. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

show 98.45% identity and 0.022 genetic distance, while chimpanzee and gorilla show 98.32% identity and 0.024 genetic distance (Supplementary Fig. 4). Additionally, our transcriptomic estimation of chimpanzee-human substitution rates closely reflects those observed genome-wide (The Chimpanzee Sequencing and Analysis Consortium, 2005). These results demonstrate the robustness of our comparative analysis protocol and show that *E. carnea* and *E. lineata* are slightly more divergent than humans and chimpanzees, suggesting that they are two distinct species.

3.3. Variant analysis reveals a common origin of E. carnea and E. sp. (parasite) populations

The analyses above suggest a very close relationship of *E. carnea* and

the parasite. However, this does not preclude two very closely related species. Single nucleotide polymorphism (SNP) analysis has been widely applied to answer population genetics questions in other model and non-model organisms (Morin et al., 2004; Stetz et al., 2016; Vendrami et al., 2016). Since demographic and geographic changes leave their signatures in the genome of a species, studies on genetic diversity with SNP genotyping have been extremely effective to trace the evolutionary history of different populations (Tishkoff et al., 2009; Campbell and Tishkoff, 2010; Choudhury et al., 2014). SNPs can serve as excellent markers to distinguish populations or closely related species. We reasoned that if our *E. carnea* and *E.* sp. (parasite) samples are representing the same species, they should share many more SNPs than the closely related species *E. lineata* or *E. carnea*. To investigate whether the similarities observed in the analyses above are reflected in the genetic variance of individuals, we called and compared the SNP locations between the transcriptomes of the species under investigation thereby looking for SNPs, which are unique or shared with at least two of the three samples. The constructed 'pseudo-genome' was used as a reference genome to call SNPs using a cascade of tools, i.e. bowtie2 (Langmead and Salzberg, 2012), SAMtools (Li et al., 2009), VarScan (Koboldt et al., 2012) and finally in-house python scripts, to filter the variants. After the filtering, we collected 14,363 SNPs (1.52/kb) in *E. carnea*, 21,251 SNPs (2.14/kb) in *E. lineata* and 7925 SNPs (1.25/kb) in *E.* sp. (parasite). No discernible correlation between called SNPs and read coverage was found, ruling out the possibility that differences in library complexity or sequencing error gave rise to differences in the number of SNPs called (Supplementary Fig. 5).

We found that there are 2061 common SNP sites between *E. carnea* and *E.* sp. (parasite), compared to only 167 shared between *E. lineata* and *E. carnea* and 107 between *E. lineata* and *E.* sp. (parasite) (Fig. 4). Thus, our analysis uncovers a considerable number of common SNP sites between the *E. carnea* and the *E.* sp. (parasite), suggesting a common origin of the populations sampled. Only a moderate number of common SNP sites were found common between *E. carnea* and *E. lineata* and between *E. lineata* and *E.* sp. (parasite), indicating a common evolutionary history shared by these two species. The common SNP sites between the species also support our earlier results, which indicate that *E. carnea* and *E.* sp. (parasite) are the same species.

3.4. The two Edwardsiella populations diverged 18.7–21.6 Mya

Next, we determined the divergence time of the two *Edwardsiella* populations by a molecular clock approach. We constructed a phylogenetic tree using seven nuclear genes (Supplementary Fig. 6) and node constraints calibration points as used in Peterson et al to estimate the divergence times (Peterson et al., 2004). Based on an estimated split of bilaterians between 555 and 642 Mya (dos Reis et al., 2015), the split between *Edwardsiella* species and *N. vectensis* dates back to 184.1–212.7 Mya, while the calculated divergence time of the *E.* carnea and *E. lineata* lies between 18.7 and 21.6 Mya, ruling out a human impact in the distribution of the two populations.

3.5. Are E. lineata and E. carnea reproductively isolated?

Biological species are commonly defined by their ability to hybridize and produce viable and fertile offspring (Queiroz, 2005). However, given the difficulty to obtain live specimens of *E. carnea* and the lack of a spawning induction protocol for *E. carnea* and *E. lineata*, such a test between *E. carnea* and *E. lineata* is not possible. However, reproductive isolation also correlates with the number of compensatory base changes (CBC) of the rRNA, in particular with the one of the internal transcribed spacer 2 (ITS2). Indeed, the presence of CBC in helix II or helix III has been correlated with the reproductive isolation between the populations and therefore CBC in the ITS2 sequence in specific locations has been proposed as a molecular barcode for species identification (Coleman 2003). An early study to test this correlation
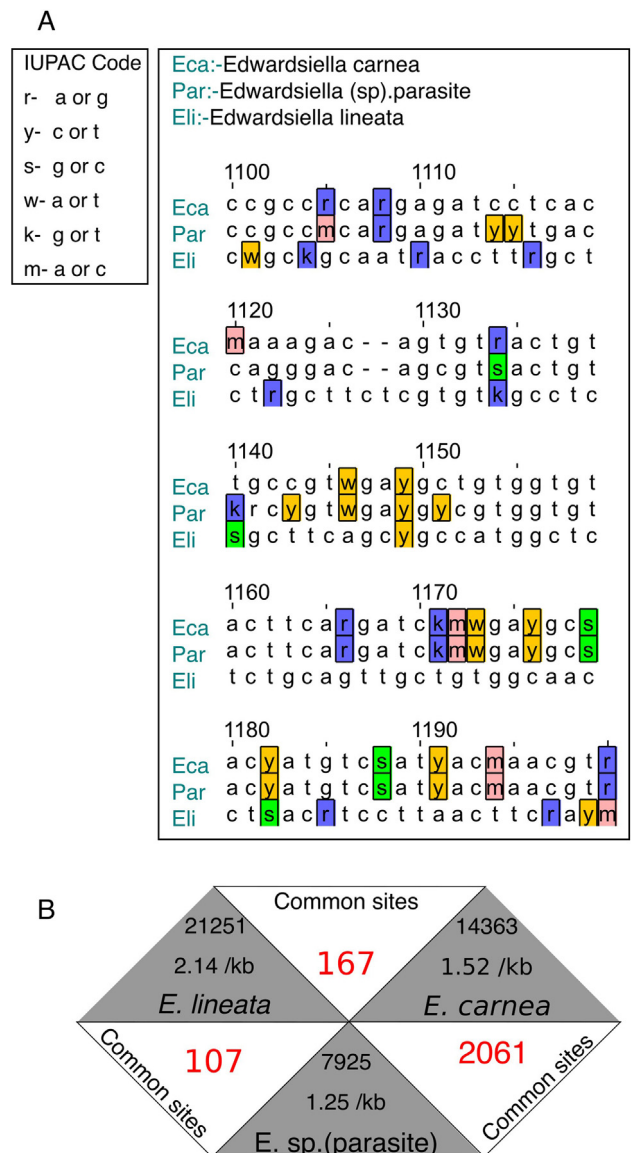

Fig. 4. Analysis of SNPs. **A:** An illustration of SNP locus comparison wherein multiple sequence alignment consisting of *E. carnea*, *E. lineata* and E. sp. (parasite) sequences with the SNP positions highlighted in IUPAC nucleotide code. SNPs were called with combination of tools such as Bowtie, SamTools, VarScan. The SNPs shown in the figure are called by the VarScan based on the read support and after applying support filters, eliminating varying nucleotides with read support less than 100 or the quality scores less than 15. Note that the highlighted nucleotides are examples of the underlying polymorphism at this position. **B:** Summary of the SNP locus comparison in 5830 orthologous sequences along with SNP per kb. There are 2061 common SNP sites between *E. carnea* and E. sp.(parasite) while as 167 common SNP sites between *E. lineata* and *E. carnea*. We found only 107 common sites between E. sp.(parasite) and *E. lineata*.

investigated 1300 closely related species and found that the presence of CBC distinguished ~93% of pairs in separate species, however, the absence of CBC could only group ~77% of pairs merged to single species (Müller et al., 2007). Another study found a clear correlation between having two or more CBCs and reproductive isolation (Pawłowska et al., 2013). Using the ITS2 sequences from the ITS2 database as reference, we identified ITS2 sequences from *N. vectensis*, *E. lineata* and *E. carnea*. We found CBCs between *N. vectensis* and *E. carnea* and between *N. vectensis* and *E. lineata* (Supplementary data 1), suggesting reproductive isolation. Notably, we did not find any CBC between *E. carnea* and *E. lineata*, raising the possibility that these two

species may still have the potential to hybridize if they were in the same environment.

3.6. Expression of Wnt genes and Wnt phylogeny

Lastly, we wished to gain insights into the genetic regulation of development by analyzing the *Wnt* genes. *Wnt* genes encode signaling molecules which are important developmental regulators involved in early axis formation, in stem cell biology and regulation of cellular differentiation processes (McMahon and Moon, 1989; Smith and Harland, 1991; Sokol et al., 1991; Christian et al., 1991; Steinbeisser et al., 1993; Wylie et al., 1996). The WNT pathway is highly conserved in all animals, but not found outside of the animal kingdom (Kusserow et al., 2005; Adamska et al., 2007). A total of 13 distinct subfamilies of WNT ligands have been characterized. While humans possess 12 of the 13 subfamilies, *Drosophila melanogaster* and *Caenorhabditis elegans* have only 6 and 3 *Wnt* genes, respectively (Prud'homme et al., 2002). Interestingly, the edwardsiid *Nematostella vectensis* expresses also 12 of the 13 known subfamilies (Kusserow et al., 2005). Hence, this sea anemone has maintained virtually all of the ancestral full complement of *Wnt* genes, which was substantially reduced in flies and nematodes. While Wnt proteins form highly conserved intra-molecular cysteine bridges, large regions of the amino acid sequence of Wnt proteins are fairly divergent, which are suitable to detect recent speciation events. Therefore, we identified the Wnt protein-coding transcripts from the transcriptomes of *E. lineata, E. carnea* and E. sp. (parasite) and constructed a phylogenetic tree (Fig. 5). As in *N. vectensis*, no Wnt9 could be detected in the investigated edwardsiids. Remarkably, although only one developmental stage was sampled in the case of *E. carnea* and *E.* sp. (parasite), almost all Wnt ligand subfamilies known from *N. vectensis* were also expressed in *E. carnea* and *E. lineata*. Only Wnt2 was missing from *E. carnea*, while Wnt7, Wnt10 and Wnt11 were missing from *E.* sp. (parasite). Generally, the Wnt protein tree corroborates the phylogenetic relationships of the *Edwardsiella* species as the 18S rRNA phylogeny. Wherever a Wnt subfamily was present in all three species/specimens, *E. carnea* and the *E.* sp. (parasite) were almost identical. However, *E. carnea/E.* sp. (parasite) were very close to *E. lineata*, with *N. vectensis* as the closest homologs.

4. Discussion

In this study, we aimed to gain genetic and phylogenetic distinction between two edwardsiid sea anemones, *Edwardsiella carnea* and *Edwardsiella lineata*, which occur at great distance from each other, along the west coast of Sweden and the east coast of North America, respectively. *E. lineata* is known to have a facultative parasitic stage between the planula larva and the polyp stage, which inhabits the gut of the American ctenophore *Mnemiopsis leidyi* (Crowell, 1976). By comparison, the life cycle of *E. carnea* has been unclear. Since *Mnemiopsis leidyi* carrying parasitic edwardsiids has been detected regularly during the last ten years in the North Sea and the Baltic Sea, this raised the question, whether this American invasive species might have drifted with the Gulf Stream to Northern Europe and acted as carriers for the parasitic cnidarians. This raised also the possibility that *E. carnea* forms a cryptic species of *E. lineata*, as a population that was established by the delivered parasites. It was therefore imperative to determine whether *E. carnea* and *E. lineata* are two distinct species and whether the parasite found in the ctenophore *M. leydi* at the Swedish coast belongs to one of them.

4.1. E. lineata and E. carnea are closely related, yet distinct species

When *E. lineata* and *E. carnea* were compared morphologically the demarcation between the species or distinctive characteristics for the species appeared weakly defined (Daly, 2002). In order to determine the phylogenetic relationship of *E. lineata* and *E. carnea,* we first carried

out a phylogenetic analysis using 18S rRNA sequences. This showed that both *E. lineata* and *E. carnea* are very closely related species and together they form the closest sister species to *Nematostella vectensis*, a widely-used model organism for comparative genomics. This result was corroborated by a global comparison of the transcriptomes assembled as 'pseudo-genomes'. Results of percentage similarity and LogDet distance measures make it evident that, although *E. lineata* and *E. carnea* are very close they are distinct on transcriptome level and share about 97% nucleotide identity.

It is clear that there is no perfect way to define a species. One way is to test reproductive isolation by crossing representatives of both populations, according to the biological species concept, although a number of counter examples are known. However, testing even enforced hybridization was not possible because obtaining the specimens of *E. carnea* is very difficult and moreover, there is no reliable spawning protocol in the lab for *E. carnea* and *E. lineata*.

Cytochrome c oxidase subunit I (COI), which was used earlier as a phylogenetic marker, shows very low sequence divergence in cnidarian species (Hebert et al., 2003). Therefore, we used the number of compensatory base changes (CBC) of the ITS2 of the rRNA as a proxy for reproductive isolation, although it is based on correlation alone (Coleman, 2007) and there are other limitations of this method such as its dependence on the secondary structure prediction algorithms (Caisová et al., 2011). Some studies suggest that 2 or more CBCs in the ITS2 sequence correlate with reproductive isolation of distinct species. Yet, while the CBC method can distinguish most animal species in over 90% of the cases, in Cnidaria only 77% of the pair-wise comparisons CBCs correlate with distinct species (Yao et al., 2010). As the calculated divergence time two *Edwardsiella* populations is roughly 20 Mio years we conclude that the two populations are indeed two distinct, yet very closely related species that are geographically isolated. Nevertheless, we do not find any CBC between *E. lineata* and *E. carnea*, raising the possibility that they could still interbreed. We conclude that these two populations are so close that the CBC method does not recognize them confidently as two separate species.

4.2. Edwardsiella carnea also has a parasitic stage, which can infect foreign ctenophores

When we included the parasite in these analyses, we unambiguously found that the parasite within the comb jelly is genetically almost identical to *E. carnea*. In the global transcriptome analysis the parasite is 99.38% identical to *E. carnea*; in the phylogenetic tree of the 18S rRNA and the *Wnt* genes, *E. carnea* is always identified as the closest relative, in many cases indistinguishable from it. This is further supported by analysis of the SNPs, where about 20 times more are shared between *E. carnea* and the parasite. Although it was not feasible to collect multiple specimens from different locations, it is likely that the minor differences between the polyps of *E. carnea* and the parasite reflect the intra-population variation. Therefore, we can rule out our initial hypothesis that the parasite stems from *E. lineata*. Thus, the parasite isolated from *M. leidyi* in Sweden is a parasitic stage of *E. carnea*. This observation is remarkable as the parasite is native to the North Sea while its host is predominantly found on the North American east coast, though periodically has been found in the North Sea. Our findings suggest that *E. carnea* is a non-selective parasite to ctenophore species as we found the parasite within *M. leidyi,* which is probably not its common host. Indeed, earlier evidence suggested it parasitized *B. infundibulum* (Stephenson, 1935; Selander et al., 2009). Interestingly, *Edwardsiella* species might have a wider host range than initially appreciated, as at least one so far unconfirmed study reported the occurrence of a parasitic stage of an edwardsiid-like organism in a scyphozoan jellyfish of the genus *Aurelia* in Croatia (Chiaverano et al., 2015).
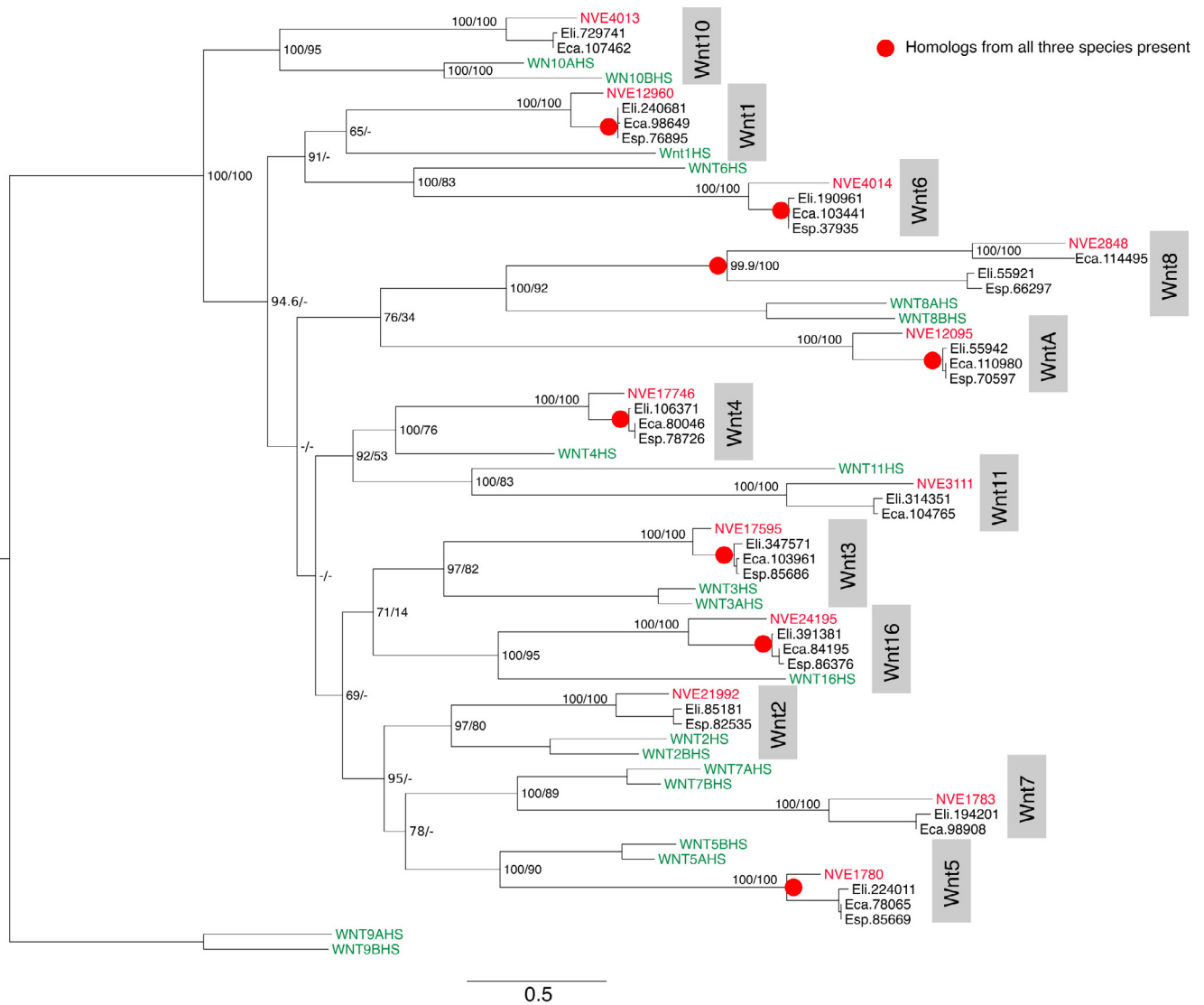
Fig. 5. Phylogenetic analysis of *Wnt* family genes of *Homo sapiens* (green), *Nematostella vectensis* (purple), *Edwardsiella lineata*, *Edwardsiella carnea*, *Edwardsiella* sp. parasite. The tree topology shown here is supported by both maximum likelihood and Bayesian inference. The ML method implemented in IQ-tree (Nguyen et al., 2015) was used. The algorithm selects the best substitution model for the alignment, which in this case was LG + I + G4. The first values shown at the nodes are percentage of 500 bootstrap runs supporting the node. The second values are the posterior probability values from the Bayesian inference. Bootstrap values below 50% and percentage posterior probability values less than 70 are not considered, indicated with "−". (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.3. Biogeography and speciation of a facultative parasitic cnidarian

Accepting that *E. lineata* and *E. carnea* form two very closely related, yet distinct species, it remains striking that such closely related species populate habitats that are almost 6000 km apart. This raises the question, how the two populations were established to give rise to two species.

There are studies estimating the range of drifting cnidarian larvae: In a recent study of coral reefs it was estimated that the dispersal range of coral larvae ranges from ~10 km to ~50 km (Markey et al., 2016). The sea anemone *Isarachnanthus nocturnus* may have a range of 2000–4000 km for the dispersal of the free swimming larvae, although the range is limited along the coast and not crossing an ocean. *I. nocturnus* also has longer larval stage of 63–118 days (Stampar et al., 2015). Since the non-parasitic form of the edwardsiid larva is non-feeding, crossing the Atlantic seems impossible for planula larvae. In the case of host-mediated transfer via *M. leidyi* the speed would be faster than the one of the larva, considering the larger size of the comb

jelly. However, if we consider factors like the ocean currents such as Gulf Stream the speed would increase substantially. The Gulf Stream with an average speed of 6.4 km/hr would make the host-mediated transport seem plausible under optimal conditions, although probably extremely rare.

However, the Gulf Stream only established after the rise of isthmus of Panama in the Cenozoic era, which caused the separation of the two oceans. While some recent studies proposed that the rise isthmus of Panama might have occurred between 7 and 23 Mya (Brady, 2017; Bacon et al., 2015; Montes et al., 2015), a more recent rise of the isthmus between 2 and 5 Mya appears to be the currently accepted view in the field (O'Dea et al., 2016). If the latter is the case, then the divergence time predates this event and we can rule out the scenario that the European population was established through the drift of the host by the Gulf Stream.

On the other hand, the supercontinent Pangaea has already split long before the divergence of the two *Edwardsiella* populations giving rise to the American, the Eurasian and the African continent (Dietz and

Holden, 1970a, 1970b). It is conceivable that in the northern territories North America and Europe remained much closer together than today, even at times of divergence of the two *Edwardsiella* species. In this scenario, it is possible that a common ancestor population that was contiguously distributed along the northern coasts of America and Europe became split by further separation of the continents. Further sampling of these species in Canada, Greenland and Northern Scandinavia would be required to evaluate the precise phylogeographic distribution of these two species.

5. Conclusions

We here show that the *Edwardsiella carnea* found at North Sea close to Sweden and Norway also has a facultative parasitic stage and is a non-selective parasite to ctenophore hosts. Our work unravels the phylogenetic relationship of two edwardsiid species, which are closest to the non-parasitic model cnidarian *Nematostella vectensis*, and the possible impact of their parasitic life cycle on the speciation events and the resulting biogeography of the species. Our analysis suggests that *Edwardsiella carnea* and *Edwardsiella lineata* are two distinct species with the possibility of crossbreeding. Moreover, we prove that the parasite found in *Mnemiopsis leidyi* is *Edwardsiella carnea*, for which no parasitic stage has been described to date.

Acknowledgements

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ympev.2018.04.035.

References

Adamska, M., Degnan, S.M., Green, K.M., Adamski, M., Craigie, A., Larroux, C., Degnan, B.M., 2007. Wnt and TGF-β Expression in the sponge Amphimedon Queenslandica and the origin of metazoan embryonic patterning. Plos One 2, e1031. http://dx.doi.org/10.1371/journal.pone.0001031.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2.

Bacon, C.D., Silvestro, D., Jaramillo, C., Smith, B.T., Chakrabarty, P., Antonelli, A., 2015. Biological evidence supports an early and complex emergence of the Isthmus of Panama. Proc. Natl. Acad. Sci. U.S.A. 112, 6110–6115. http://dx.doi.org/10.1073/pnas.1423853112.

Bateman, A., Agrawal, S., Birney, E., Bruford, E.A., Bujnicki, J.M., Cochrane, G., Cole, J.R., Dinger, M.E., Enright, A.J., Gardner, P.P., Gautheret, D., Griffiths-Jones, S., Harrow, J., Herrero, J., Holmes, I.H., Huang, H.-D., Kelly, K.A., Kersey, P., Kozomara, A., Lowe, T.M., Marz, M., Moxon, S., Pruitt, K.D., Samuelsson, T., Stadler, P.F., Vilella, A.J., Vogel, J.-H., Williams, K.P., Wright, M.W., Zwieb, C., 2011. RNAcentral: a vision for an international database of RNA sequences. RNA 17, 1941–1946. http://dx.doi.org/10.1261/rna.2750811.

Boersma, M., Malzahn, A.M., Greve, W., Javidpour, J., 2007. The first occurrence of the ctenophore Mnemiopsis leidyi in the North Sea. Helgol. Mar. Res. 61, 153. http://dx.doi.org/10.1007/s10152-006-0055-2.

Bolger, et al., 2014. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics.

Brady, S.G., 2017. Army ant invasions reveal phylogeographic processes across the Isthmus of Panama. Mol. Ecol. 26, 703–705. http://dx.doi.org/10.1111/mec.13981.

Bridge, D., Cunningham, C.W., DeSalle, R., Buss, L.W., 1995. Class-level relationships in the phylum Cnidaria: molecular and morphological evidence. Mol. Biol. Evol. 12, 679–689. http://dx.doi.org/10.1093/oxfordjournals.molbev.a040246.

Bridge, D., Cunningham, C.W., Schierwater, B., DeSalle, R., Buss, L.W., 1992. Class-level relationships in the phylum Cnidaria: evidence from mitochondrial genome structure. PNAS 89, 8750–8753. http://dx.doi.org/10.1073/pnas.89.18.8750.

Bumann, D., Puls, G., 1996. Infestation with larvae of the sea anemone Edwardsia lineata affects nutrition and growth of the ctenophore Mnemiopsis leidyi. Parasitology 113, 123–128. http://dx.doi.org/10.1017/S0031182000066361.

Cai, T.T., Liang, T., Zhou, H.H., 2015. Law of log determinant of sample covariance matrix and optimal estimation of differential entropy for high-dimensional Gaussian distributions. J. Multivariate Anal. 137, 161–172. http://dx.doi.org/10.1016/j.jmva.2015.02.003.

Caisová, L., Marin, B., Melkonian, M., 2011. A close-up view on ITS2 evolution and speciation – a case study in the Ulvophyceae (Chlorophyta, Viridiplantae). BMC Evol. Biol. 11, 262. http://dx.doi.org/10.1186/1471-2148-11-262.

Campbell, M.C., Tishkoff, S.A., 2010. The evolution of human genetic and phenotypic variation in Africa. Curr. Biol. 20, R166–R173. http://dx.doi.org/10.1016/j.cub.2009.11.050.

Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973. http://dx.doi.org/10.1093/bioinformatics/btp348.

Chiaverano, L.M., Graham, W.M., Costello, J.H., 2015. Parasites alter behavior, reproductive output, and growth patterns of Aurelia medusae in a marine lake. Mar. Ecol. Prog. Ser. 540, 87–98. http://dx.doi.org/10.3354/meps11513.

Choudhury, A., Hazelhurst, S., Meintjes, A., Achinike-Oduaran, O., Aron, S., Gamieldien, J., Jalali Sefid Dashti, M., Mulder, N., Tiffin, N., Ramsay, M., 2014. Population-specific common SNPs reflect demographic histories and highlight regions of genomic plasticity with functional relevance. BMC Genomics 15, 437. http://dx.doi.org/10.1186/1471-2164-15-437.

Christian, J.L., McMahon, J.A., McMahon, A.P., Moon, R.T., 1991. Xwnt-8, a Xenopus Wnt-1/int-1-related gene responsive to mesoderm-inducing growth factors, may play a role in ventral mesodermal patterning during embryogenesis. Development 111, 1045–1055.

Coleman, A.W., 2007. Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. Nucl. Acids Res. 35, 3322–3329. http://dx.doi.org/10.1093/nar/gkm233.

Crowell, S., 1976. An Edwardsiid Larva Parasitic in Mnemiopsis. In: Coelenterate Ecology and Behavior. Springer, Boston, MA, pp. 247–250. https://doi.org/10.1007/978-1-4757-9724-4_26.

Daly, M., 2002. A systematic revision of Edwardsiidae (Cnidaria, Anthozoa). Invertebr. Biol. 121. http://dx.doi.org/10.1111/j.1744-7410.2002.tb00061.x.

Daly, M., Lipscomb, D.L., Allard, M.W., 2002. A simple test: evaluating explanations for the relative simplicity of the Edwardsiidae (cnidaria: Anthozoa). Evolution 56, 502–510. http://dx.doi.org/10.1111/j.0014-3820.2002.tb01361.x.

Darling, J.A., Reitzel, A.R., Burton, P.M., Mazza, M.E., Ryan, J.F., Sullivan, J.C., Finnerty, J.R., 2005. Rising starlet: the starlet sea anemone, Nematostella vectensis. Bioessays 27, 211–221. http://dx.doi.org/10.1002/bies.20181.

Database Resources of the National Center for Biotechnology Information, 2017. Nucleic Acids Res 45, D12–D17. https://doi.org/10.1093/nar/gkw1071.

Dietz, R.S., Holden, J.C., 1970a. The breakup of pangaea. Sci. Am. 223, 30–41.

Dietz, R.S., Holden, J.C., 1970b. Reconstruction of Pangaea: breakup and dispersion of continents, Permian to Present. J. Geophys. Res. 75, 4939–4956. http://dx.doi.org/10.1029/JB075i026p04939.

dos Reis, M., Thawornwattana, Y., Angelis, K., Telford, M.J., Donoghue, P.C.J., Yang, Z., 2015. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. Curr. Biol. 25, 2939–2950. http://dx.doi.org/10.1016/j.cub.2015.09.066.

Ereskovsky, A.V., Richter, D.J., Lavrov, D.V., Schippers, K.J., Nichols, S.A., 2017. Transcriptome sequencing and delimitation of sympatric Oscarella species (O. carmela and O. pearsei sp. nov) from California, USA. PLOS ONE 12, e0183002. http://dx.doi.org/10.1371/journal.pone.0183002.

Faasse, M.A., Bayha, K.M., 2006. The ctenophore Mnemiopsis leidyi A. Agassiz 1865 in coastal waters of the Netherlands: an unrecognized invasion? Aquat. Invasions. http://dx.doi.org/10.3391/ai.2006.1.4.10.

Genikhovich, G., Technau, U., 2009. The starlet sea anemone Nematostella vectensis: an anthozoan model organism for studies in comparative genomics and functional evolutionary developmental biology. Cold Spring Harb Protoc 2009, pdb.emo129. https://doi.org/10.1101/pdb.emo129.

Gosse, P.H., 1856. A manual of marine zoology for the British Isles. J. Van Voorst.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652. http://dx.doi.org/10.1038/nbt.1883.

Haas, B., Papanicolaou, A., 2012. TransDecoder (Find Coding Regions within Transcripts).

Hand, C., 1990. Controlled Culture of a New Marine Model, the Sea Anemone Nematostella Vectensis. Biennial Report of Completed Projects 70.

Hand, C., Uhlinger, K.R., 1992. The culture, sexual and asexual reproduction, and growth of the sea anemone nematostella vectensis. Biol. Bull. 182, 169–176. http://dx.doi.org/10.2307/1542110.

Hansson, Hans G., 2006. Ctenophores of the Baltic and adjacent Seas – the invader Mnemiopsis is here! – Mnemiopsis. HGH.pdf. Aquatic Invasions. vol. 1, pp. 295–298.

Harris, R.S., 2007. Improved pairwise alignment of genomic DNA. Ph.D. Thesis, The Pennsylvania State University.

Hebert, P.D.N., Ratnasingham, S., de Waard, J.R., 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc. R. Soc. London B: Biol. Sci. 270, S96–S99. http://dx.doi.org/10.1098/rsbl.2003.0025.

Javidpour, J., Sommer, U., Shiganova, T., 2006. First record of Mnemiopsis leidyi A. Agassiz 1865 in the Baltic Sea. Aquatic Invasions 1, pp. 299–302. https://doi.org/10.3391/ai.2006.1.4.17.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. http://dx.doi.org/10.1093/molbev/mst010.

Kayal, E., Roure, B., Philippe, H., Collins, A.G., Lavrov, D.V., 2013. Cnidarian phylogenetic relationships as revealed by mitogenomics. BMC Evol. Biol. 13, 5. http://dx.doi.org/10.1186/1471-2148-13-5.

Kinsella, R.J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., Kersey, P., Flicek, P., 2011. Ensembl

BioMarts: a hub for data retrieval across taxonomic space. Database (Oxford) 2011. http://dx.doi.org/10.1093/database/bar030.

Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., Wilson, R.K., 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 22, 568–576. http://dx.doi.org/10.1101/gr.129684.111.

Kusserow, A., Pang, K., Sturm, C., Hrouda, M., Lentfer, J., Schmidt, H.A., Technau, U., von Haeseler, A., Hobmayer, B., Martindale, M.Q., Holstein, T.W., 2005. Unexpected complexity of the Wnt gene family in a sea anemone. Nature 433, 156–160. http://dx.doi.org/10.1038/nature03158.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Meth. 9, 357–359. http://dx.doi.org/10.1038/nmeth.1923.

Layden, M.J., Johnston, H., Amiel, A.R., Havrilak, J., Steinworth, B., Chock, T., Röttinger, E., Martindale, M.Q., 2016a. MAPK signaling is necessary for neurogenesis in Nematostella vectensis. BMC Biol. 14, 61. http://dx.doi.org/10.1186/s12915-016-0282-1.

Layden, M.J., Rentzsch, F., Röttinger, E., 2016b. The rise of the starlet sea anemone Nematostella vectensis as a model system to investigate development and regeneration. WIREs Dev. Biol. 5, 408–428. http://dx.doi.org/10.1002/wdev.222.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079. http://dx.doi.org/10.1093/bioinformatics/btp352.

Li, L., Stoeckert, C.J., Roos, D.S., 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13, 2178–2189. http://dx.doi.org/10.1101/gr.1224503.

Lorenz, R., Bernhart, S.H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., Hofacker, I.L., 2011. ViennaRNA Package 2.0. Algorithms for Molecular Biology 6, 26. https://doi.org/10.1186/1748-7188-6-26.

Markey, K.L., Abdo, D.A., Evans, S.N., Bosserelle, C., 2016. Keeping it local: dispersal limitations of coral larvae to the high latitude coral reefs of the Houtman Abrolhos Islands. PLOS ONE 11, e0147628. http://dx.doi.org/10.1371/journal.pone.0147628.

Massingham, T., Goldman, N., 2007. Statistics of the log-det estimator. Mol. Biol. Evol. 24, 2277–2285. http://dx.doi.org/10.1093/molbev/msm160.

McMahon, A.P., Moon, R.T., 1989. Ectopic expression of the proto-oncogene int-1 in Xenopus embryos leads to duplication of the embryonic axis. Cell 58, 1075–1084. http://dx.doi.org/10.1016/0092-8674(89)90506-0.

Merget, B., Koetschan, C., Hackl, T., Förster, F., Dandekar, T., Müller, T., Schultz, J., Wolf, M., 2012. The ITS2 database. J. Vis. Exp. http://dx.doi.org/10.3791/3806.

Montes, C., Cardona, A., Jaramillo, C., Pardo, A., Silva, J.C., Valencia, V., Ayala, C., Pérez-Angel, L.C., Rodriguez-Parra, L.A., Ramirez, V., Niño, H., 2015. Middle Miocene closure of the Central American Seaway. Science 348, 226–229. http://dx.doi.org/10.1126/science.aaa2815.

Morin, P.A., Luikart, G., Wayne, R.K., group the S. workshop, 2004. SNPs in ecology, evolution and conservation. Trends Ecol. Evol. 19, 208–216. http://dx.doi.org/10.1016/j.tree.2004.01.009.

Müller, T., Philippi, N., Dandekar, T., Schultz, J., Wolf, M., 2007. Distinguishing species. RNA 13, 1469–1472. http://dx.doi.org/10.1261/rna.617107.

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274. http://dx.doi.org/10.1093/molbev/msu300.

Nyholm, K.G., 1943. Zur Entwicklung und Entwicklungsbiologie der Ceriantharien und Aktinien. Zool. Bidr Upps 22, 87–248.

O'Dea, A., Lessios, H.A., Coates, A.G., Eytan, R.I., Restrepo-Moreno, S.A., Cione, A.L., Collins, L.S., de Queiroz, A., Farris, D.W., Norris, R.D., Stallard, R.F., Woodburne, M.O., Aguilera, O., Aubry, M.-P., Berggren, W.A., Budd, A.F., Cozzuol, M.A., Coppard, S.E., Duque-Caro, H., Finnegan, S., Gasparini, G.M., Grossman, E.L., Johnson, K.G., Keigwin, L.D., Knowlton, N., Leigh, E.G., Leonard-Pingel, J.S., Marko, P.B., Pyenson, N.D., Rachello-Dolmen, P.G., Soibelzon, E., Soibelzon, L., Todd, J.A., Vermeij, G.J., Jackson, J.B.C., 2016. Formation of the isthmus of panama. Sci. Adv. 2, e1600883. http://dx.doi.org/10.1126/sciadv.1600883.

Pawłowska, J., Walther, G., Wilk, M., de Hoog, S., Wrzosek, M., 2013. The use of compensatory base change analysis of ITS2 as a tool in the phylogeny of Mucorales, illustrated by the Mucor circinelloides complex. Org. Divers. Evol. 13, 497–502. http://dx.doi.org/10.1007/s13127-013-0139-1.

Peterson, K.J., Lyons, J.B., Nowak, K.S., Takacs, C.M., Wargo, M.J., McPeek, M.A., 2004. Estimating metazoan divergence times with a molecular clock. PNAS 101, 6536–6541. http://dx.doi.org/10.1073/pnas.0401670101.

Prud'homme, B., Lartillot, N., Balavoine, G., Adoutte, A., Vervoort, M., 2002. Phylogenetic analysis of the Wnt gene family. Curr. Biol. 12, 1395–1400. http://dx.doi.org/10.1016/S0960-9822(02)01068-0.

de Queiroz, K., 2005. Ernst Mayr and the modern concept of species. PNAS 102, 6600–6607. http://dx.doi.org/10.1073/pnas.0502030102.

Reitzel, A.M., Daly, M., Sullivan, J.C., Finnerty, J.R., 2009. Comparative anatomy and histology of developmental and parasitic stages in the life cycle of the lined sea anemone Edwardsiella lineata. J. Parasitol. 95, 100–112. http://dx.doi.org/10.1645/GE-1623.1.

Reitzel, A.M., Sullivan, J.C., Brown, B.K., Chin, D.W., Cira, E.K., Edquist, S.K., Genco, B.M., Joseph, O.C., Kaufman, C.A., Kovitvongsa, K., Muñoz, M.M., Negri, T.L., Taffel, J.R., Zuehlke, R.T., Finnerty, J.R., 2007. Ecological and developmental dynamics of a

host-parasite system involving a sea anemone and two ctenophores. J. Parasitol. 93, 1392–1402. http://dx.doi.org/10.1645/GE-1250.1.

Reitzel, A.M., Sullivan, J.C., Finnerty, J.R., 2006. Qualitative shift to indirect development in the parasitic sea anemone Edwardsiella lineata. Oxford J. 46, 827–837.

Rentzsch, F., Technau, U., 2016. Genomics and development of Nematostella vectensis and other anthozoans. Curr. Opin. Genet. Dev. 39, 63–70. http://dx.doi.org/10.1016/j.gde.2016.05.024.

Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: the European molecular biology open software suite. Trends Genet. 16, 276–277. http://dx.doi.org/10.1016/S0168-9525(00)02024-2.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Oxford J. Bioinf. 19, 1572–1574.

Sanderson, M.J., 2003. r8s: Inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics (Oxford, England) 19, pp. 301–302. https://doi.org/10.1093/bioinformatics/19.2.301.

Seibel, P.N., Müller, T., Dandekar, T., Wolf, M., 2008. Synchronous visual analysis and editing of RNA sequence and secondary structure alignments using 4SALE. BMC Res. Notes 1, 1–7. http://dx.doi.org/10.1186/1756-0500-1-91.

Selander, E., Møller, L.F., Sundberg, P., Tiselius, P., 2009. Parasitic anemone infects the invasive ctenophore Mnemiopsis leidyi in the North East Atlantic. Biol Invasions 12, 1003–1009. http://dx.doi.org/10.1007/s10530-009-9552-y.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics btv351. https://doi.org/10.1093/bioinformatics/btv351.

Smith, W.C., Harland, R.M., 1991. Injected Xwnt-8 RNA acts early in Xenopus embryos to promote formation of a vegetal dorsalizing center. Cell 67, 753–765. http://dx.doi.org/10.1016/0092-8674(91)90070-F.

Sokol, S., Christian, J.L., Moon, R.T., Melton, D.A., 1991. Injected Wnt RNA induces a complete body axis in Xenopus embryos. Cell 67, 741–752. http://dx.doi.org/10.1016/0092-8674(91)90069-B.

Stampar, S.N., Morandini, A.C., Branco, L.C., da Silveira, F.L., Migotto, A.E., 2015. Drifting in the oceans: Isarachnanthus nocturnus (Cnidaria, Ceriantharia, Arachnactidae), an anthozoan with an extended planktonic stage. Mar. Biol. 162, 2161–2169. http://dx.doi.org/10.1007/s00227-015-2747-0.

Steele, R.E., David, C.N., Technau, U., 2011. A genomic view of 500 million years of cnidarian evolution. Trends Genet. 27, 7–13. http://dx.doi.org/10.1016/j.tig.2010.10.002.

Stefanik, D.J., Lubinski, T.J., Granger, B.R., Byrd, A.L., Reitzel, A.M., DeFilippo, L., Lorenc, A., Finnerty, J.R., 2014. Production of a reference transcriptome and transcriptomic database (EdwardsiellaBase) for the lined sea anemone, Edwardsiella lineata, a parasitic cnidarian. BMC Genomics 15, 71. http://dx.doi.org/10.1186/1471-2164-15-71.

Steinbeisser, H., Robertis, E.M.D., Ku, M., Kessler, D.S., Melton, D.A., 1993. Xenopus axis formation: induction of goosecoid by injected Xwnt-8 and activin mRNAs. Development 118, 499–507.

Stephenson, T., 1935. The British Sea Anemones. Ray Society Monograph No. 121. London, vol. 2, p. 426.

Stetz, J.B., smith, S., Sawaya, M.A., Ramsey, A.B., Amish, S.J., Schwartz, M.K., Luikart, G., 2016. Discovery of 20,000 RAD–SNPs and development of a 52-SNP array for monitoring river otters. Conserv. Genet. Resour. 8, 299–302. http://dx.doi.org/10.1007/s12686-016-0558-3.

Technau, U., Schwaiger, M., 2015. Recent advances in genomics and transcriptomics of cnidarians. Mar. Genom., Mar. Genom. Evol. Dev. 24, 131–138. http://dx.doi.org/10.1016/j.margen.2015.09.007.

The Chimpanzee Sequencing and Analysis Consortium, Waterson, R.H., Lander, E.S., Wilson, R.K., 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437, 69–87. http://dx.doi.org/10.1038/nature04072.

Tishkoff, S.A., Reed, F.A., Friedlaender, F.R., Ehret, C., Ranciaro, A., Froment, A., Hirbo, J.B., Awomoyi, A.A., Bodo, J.-M., Doumbo, O., Ibrahim, M., Juma, A.T., Kotze, M.J., Lema, G., Moore, J.H., Mortensen, H., Nyambo, T.B., Omar, S.A., Powell, K., Pretorius, G.S., Smith, M.W., Thera, M.A., Wambebe, C., Weber, J.L., Williams, S.M., 2009. The genetic structure and history of Africans and African Americans. Science 324, 1035–1044. http://dx.doi.org/10.1126/science.1172257.

Vendrami, D.L.J., Shah, A., Telesca, L., Hoffman, J.I., 2016. Mining the transcriptomes of four commercially important shellfish species for single nucleotide polymorphisms within biomineralization genes. Mar. Genom. 27, 17–23. http://dx.doi.org/10.1016/j.margen.2015.12.009.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J., 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25, 1189–1191. http://dx.doi.org/10.1093/bioinformatics/btp033.

Wylie, C., Kofron, M., Payne, C., Anderson, R., Hosobuchi, M., Joseph, E., Heasman, J., 1996. Maternal beta-catenin establishes a 'dorsal signal' in early Xenopus embryos. Development 122, 2987–2996.

Yao, H., Song, J., Liu, C., Luo, K., Han, J., Li, Y., Pang, X., Xu, H., Zhu, Y., Xiao, P., Chen, S., 2010. Use of ITS2 Region as the Universal DNA barcode for plants and animals. PLOS ONE 5, e13102. http://dx.doi.org/10.1371/journal.pone.0013102.

Zheng, W., Chung, L.M., Zhao, H., 2011. Bias detection and correction in RNA-Sequencing data. BMC Bioinf. 12, 290. http://dx.doi.org/10.1186/1471-2105-12-290.