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

Published in:
Molecular Phylogenetics and Evolution

Link to article, DOI:
10.1016/j.ympev.2018.04.035

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Dispersal and speciation: The cross Atlantic relationship of two parasitic cnidarians

Rohit Dnyansagar\(^a,1\), Bob Zimmermann\(^b,1\), 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

A B S T R A C T

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 \(E.\) 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. 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...
Darling et al., 2005; Genikhovich and Technau, 2009; Layden et al., 2016; Rentzsch and Technau, 2016; Steele et al., 2011; Technau and Schwaiger, 2015) \textit{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 \textit{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 \textit{Nematostella}, are \textit{Edwardsiella lineata} and \textit{Edwardsiella carnea} (Fig. 1). \textit{E. lineata} is found at the east coast of North America. Interestingly, planula larvae of \textit{E. lineata} can enter the gut of ctenophores, mostly \textit{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 \textit{M. leidyi}, \textit{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). \textit{E. carnea}, which is found along the Atlantic coast of Sweden and Norway (Gosse, 1856; Daly, 2002) is extremely similar morphologically to \textit{E. lineata}. Indeed, they differ only by two out of sixty morphological characters (Daly, 2002), raising the question of their phylogenetic relationship. \textit{E. carnea} was first identified as \textit{Edwardsia carnea} (Gosse, 1856). Since then the species has been renamed several times as \textit{Halcampa microps} (Gosse, 1856), \textit{Milneedwardsia carnea} (Carlgren, 1921) and \textit{Fagesia carnea} (Delphy, 1938). The name \textit{Edwardsiella carnea} is currently accepted by \textit{The World Register for Marine Species}, which is based on \textit{The Marine Fauna of the British Isles and North-West Europe}. Although \textit{E. carnea} was 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 \textit{E. lineata}, parasitic stages were never described for \textit{E. carnea}.

The American ctenophore \textit{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 \textit{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 verimform 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 GGGD000000000 and GGBB000000000 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 ‘−notrim’ 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 gibertensis, Edwardsia andrei, Edwardsia japonica, Edwardsia elegans, Edwardsia sipunculoides, Edwardsia timida, Edwardsia tuberculata and Nematostella
best alignment was edited manually using Jalview (Waterhouse (parasite) transcriptomes, 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 tblastn 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 the 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 XFast format (fasta sequence with secondary structure) was prepared and used in 4SALE program (Seibol 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 bisphosphate, 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). Alignments of individual genes were performed with MAFFT in E-INS-i mode (Katoh and Standley, 2013) and trimmed using trimAl in automated1 mode (Capella-Gutierrez 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 bilateralian 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 Nemastellata vectensis, we collected four specimens of the Edwardsiid sea anemone E. carnea on the coast of Lysekil, Sweden. Likewise, we pooled twelve specimens of parasites from the cnemophore 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 (MrBays 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 data 1).
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. 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.)
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 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 ITIS2 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 (Kussrow 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 (Kussrow 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. leidyi 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 lineata 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).
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 $\sim 10$ km to $\sim 50$ 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 contiguous 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 Scandi- navia 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 Edwardsiellid 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

This work was supported by an grant of the Austrian Science Fund FWF to U.T. (P24858).

Appendix A. Supplementary data

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

References
