Extreme Mitogenomic Variation in Natural Populations of Chaetognaths
Marlétag, F.; Le Parco, Y.; Liu, S.; Peijnenburg, K.T.C.A.

Published in:
Genome Biology and Evolution

DOI:
10.1093/gbe/evx090

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Extreme Mitogenomic Variation in Natural Populations of Chaetognaths

Ferdinand Marléťaz1,2,*, Yannick Le Parco3, Shenglin Liu4, and Katja TCA Peijnenburg4,5,*

1Department of Zoology, University of Oxford, United Kingdom
2Molecular Genetics Unit, Okinawa Institute of Science and Technology, Onna, Japan
3Institut Méditerranéen d’Océanologie (CNRS UMR 7294), Aix-Marseille Université, Campus de Luminy, Marseille, France
4Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, The Netherlands
5Naturalis Biodiversity Center, Marine Biodiversity, Leiden, The Netherlands

*Corresponding authors: E-mails: ferdinand.marletaz@gmail.com; katja.peijnenburg@naturalis.nl.

Accepted: May 9, 2017

Data deposition: Assembled mitochondrial genomes have been deposited in Genbank under the accessions KP899748-KP899801 (details in Table S1). Sequences of genotyped markers are available under the accessions: KP843748-KP843841 and KP857119-KP857568 (details in Table S1). Transcriptome data was deposited with the accession PRJNA357934

Abstract

The extent of within-species genetic variation across the diversity of animal life is an underexplored problem in ecology and evolution. Although neutral genetic variation should scale positively with population size, mitochondrial diversity levels are believed to show little variation across animal species. Here, we report an unprecedented case of extreme mitochondrial diversity within natural populations of two morphospecies of chaetognaths (arrow worms). We determine that this diversity is composed of deep sympatric mitochondrial lineages, which are in some cases as divergent as human and platypus. Additionally, based on 54 complete mitogenomes, we observed mitochondrial gene order differences between several of these lineages. We examined nuclear divergence patterns (18S, 28S, and an intron) to determine the possible origin of these lineages, but did not find congruent patterns between mitochondrial and nuclear markers. We also show that extreme mitochondrial divergence in chaetognaths is not driven by positive selection. Hence, we propose that the extreme levels of mitochondrial variation could be the result of either a complex scenario of reproductive isolation, or a combination of large population size and accelerated mitochondrial mutation rate. These findings emphasize the importance of characterizing genome-wide levels of nuclear variation in these species and promote chaetognaths as a remarkable model to study mitochondrial evolution.

Key words: genetic diversity, mitochondrial genomes, molecular evolution, marine invertebrates, Chaetognatha.

Introduction

Genetic diversity estimates provide information about a species’ population history, dynamics, and adaptive potential (Ellegren and Galtier 2016; Leffler et al. 2012; Peijnenburg and Goetze 2013; Reed and Frankham 2003). Mitochondrial genes have been widely used to characterize natural populations at the molecular level, mostly because of technical ease-of-use considerations such as clonality and high mutation rate (Galtier et al. 2009). In particular, the DNA barcoding approach has stimulated the use of mitochondrial gene fragments, such as Cytochrome Oxidase 1 (cox1). The combination of strong interspecific divergence and generally low levels of intraspecific variation, also known as the barcoding “gap”, makes such markers an essential tool for species identification (Bucklin et al. 2011). Many observations of high levels of mitochondrial divergence have consequently been interpreted as evidence of cryptic speciation, the distinct genetic entities being often, but not always, associated with previously overlooked geographic, morphological, or reproductive boundaries (Bickford et al. 2007).

The observation that the range of genetic variation does not scale with population size as predicted by the neutral theory of molecular evolution was referred to by Lewontin as the “paradox of variation” (Hahn 2008; Lewontin 1974).
This paradox is particularly manifest for animal mitochondrial diversity levels, which show surprisingly little variation across species (Bazin et al. 2006). This has been interpreted as the consequence of pervasive positive selection (Bazin et al. 2006), of genetic hitch-hiking (Gillespie 2001), or as the result of an inverse correlation between population size and mutation rate (Piganeau and Eyre-Walker 2009). Similarly, mitochondrial transplantation experiments have demonstrated that the cyto-nuclear interactions in respiratory complexes are an important factor limiting diversity in mitochondrial genes (Kenyon and Moraes 1997; Ellison and Burton 2008). These explanations may account for the fact that the combination of high genetic diversity and signatures of strong purifying selection has rarely been observed in nature for mitochondria, including in species with large population sizes (Meiklejohn et al. 2007).

The genetics and diversity of oceanic invertebrates remain poorly characterized (Bucklin et al. 2011; GIGA Community of Scientists et al. 2014; Peijnenburg and Goetze 2013). Chaetognaths are an enigmatic marine invertebrate phylum, which possess a unique combination of morphological, developmental, and genomic characters (Telford 2004). As a possible sister group to other protostomes, chaetognaths occupy a key phylogenetic position to understand the evolution of bilaterian animals (Marlé taz et al. 2006, 2008; Matus et al. 2006). They also represent a major planktonic group and play important roles in marine food webs as the primary predators (Vannier et al. 2007), which suggests that the phylum originated during the “cambrian explosion” and underwent little morphological evolution during the course of animal evolution (Knoll and Carroll 1999). However, the phylum has a poor recent fossil record, which prevents dating divergences of the main extant clades and its taxonomy is still unresolved (Vannier et al. 2006; Gasm i et al. 2014). Transcriptomic studies revealed that chaetognath genomes underwent extensive diversity levels, which show surprisingly little variation across species (Bazin et al. 2006). This has been interpreted as the consequence of pervasive positive selection (Bazin et al. 2006), of genetic hitch-hiking (Gillespie 2001), or as the result of an inverse correlation between population size and mutation rate (Piganeau and Eyre-Walker 2009). Similarly, mitochondrial transplantation experiments have demonstrated that the cyto-nuclear interactions in respiratory complexes are an important factor limiting diversity in mitochondrial genes (Kenyon and Moraes 1997; Ellison and Burton 2008). These explanations may account for the fact that the combination of high genetic diversity and signatures of strong purifying selection has rarely been observed in nature for mitochondria, including in species with large population sizes (Meiklejohn et al. 2007).

Materials and Methods

Animal Sampling and Genotyping

Individuals of the species Spadella cephaloptera, were collected from Calanque de Sormiou near Marseille (France) at shallow depth on Posidonia seagrass using a plankton net (supplementary fig. S1, Supplementary Material online). Individuals of the species Sagitta elegans and S. setosa were collected using a multinet on a single daytrip in the Gullmar fjord near Kristineberg (Lye sekil, Sweden). All chaetognaths were identified under a microscope while still alive by taxonomic experts (FM and KTCAP). Only mature individuals without out food in their guts and without visible parasites were preserved and used for genetic analysis.

Genomic DNA was extracted using QiaAmp Micro Kit and DNeasy Blood & Tissue Kit (Qiagen) and used as a template for PCR amplification using GoTaq (Promega) or Phire Hot Start II DNA Polymerase (Finzymes). Number of individuals for species and each amplified marker are specified in table 1 and supplementary table S1, Supplementary Material online. Nuclear markers (18S, 28S, nuclear intron of the ribosomal protein L36a) were amplified in the same individuals for which the entire mitogenome was sequenced. Primers and protocols used for 18S and 28S characterization are described in (Papillon et al. 2006). Nuclear loci were cloned in pGEM-T (Promega) and 2–8 clones were picked and sequenced in both directions. All primers employed are specified in supplementary table S5, Supplementary Material online. All sequences were deposited in Genbank (accessions in supplementary table S1, Supplementary Material online).

Sequencing of Individual Mitochondrial Genomes

In S. cephaloptera, two half-genome fragments (6–8 kb) were amplified using specific primers located in 16S and cox1 genes using the Accuprime Taq (Invitrogen) for five individuals. These fragments were purified using S.N.A.P. gel purification kit (Invitrogen) and subsequently cloned with Topo XL cloning Kit (Invitrogen). Plasmid DNA was prepared with Plasmid Midi Kit (Qiagen). These templates were sequenced using a primer walking strategy using 6–10 Sanger reads in total.

In 37 S. elegans and 12 S. setosa individuals, outward pointing primers were designed within the cox1 and cox2 genes and used to amplify the mitogenome as a single amplicon using Phusion DNA polymerase (Thermo scientific). For eight individuals long-range PCR fragments of both cox1 and cox2 primer sets were sequenced to verify that the same mitochondrial genome was obtained. A unique library was
Table 1
Genetic Diversity Estimates for Selected Loci in Single Populations of the Chaetognaths Spadella cephaloptera, Sagitta elegans, and S. setosa

<table>
<thead>
<tr>
<th>Species/Location</th>
<th>Gene</th>
<th>n</th>
<th>K</th>
<th>π</th>
<th>Δθw</th>
<th>Δt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spadella cephaloptera</td>
<td>cox1</td>
<td>25</td>
<td>25</td>
<td>0.1385</td>
<td>0.0539</td>
<td>0.1223</td>
</tr>
<tr>
<td>Calanque of Sormiou, France</td>
<td>All genes</td>
<td>30</td>
<td>30</td>
<td>0.1181</td>
<td>0.1149</td>
<td>0.1066</td>
</tr>
<tr>
<td>Sormiou, France</td>
<td>All genes</td>
<td>5</td>
<td>5</td>
<td>0.2939</td>
<td>1.4414</td>
<td>0.2470</td>
</tr>
<tr>
<td>185s</td>
<td>8</td>
<td>8</td>
<td>0.0095</td>
<td>0.0362</td>
<td>-0.3362</td>
<td></td>
</tr>
<tr>
<td>L36a intron</td>
<td>30</td>
<td>30</td>
<td>0.0411</td>
<td>-0.2605</td>
<td>0.0443</td>
<td></td>
</tr>
<tr>
<td>Sagitta elegans</td>
<td>cox1</td>
<td>107</td>
<td>96</td>
<td>0.1775</td>
<td>3.6069</td>
<td>0.0854</td>
</tr>
<tr>
<td>Gullmar fjord, Sweden</td>
<td>All genes</td>
<td>37</td>
<td>37</td>
<td>0.2081</td>
<td>2.1922</td>
<td>0.1304</td>
</tr>
<tr>
<td>185s</td>
<td>24</td>
<td>24</td>
<td>0.0000</td>
<td>NA</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>285s</td>
<td>27</td>
<td>27</td>
<td>0.0000</td>
<td>NA</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>L36a intron</td>
<td>37</td>
<td>37</td>
<td>0.0275</td>
<td>-1.8220</td>
<td>0.0590</td>
<td></td>
</tr>
<tr>
<td>Sagitta setosa</td>
<td>cox1</td>
<td>54</td>
<td>49</td>
<td>0.0088</td>
<td>-2.0446</td>
<td>0.0214</td>
</tr>
<tr>
<td>Gullmar fjord, Sweden</td>
<td>All genes</td>
<td>12</td>
<td>12</td>
<td>0.0088</td>
<td>-1.6561</td>
<td>0.0139</td>
</tr>
<tr>
<td>185s</td>
<td>20</td>
<td>20</td>
<td>0.0001</td>
<td>-1.7800</td>
<td>0.0002</td>
<td></td>
</tr>
<tr>
<td>285s</td>
<td>20</td>
<td>20</td>
<td>0.0006</td>
<td>-1.7800</td>
<td>0.0015</td>
<td></td>
</tr>
</tbody>
</table>

Build for each individual mitogenome using the Nextera Kit (Illumina) from 30–70 ng of PCR product. Multiplexed libraries were then sequenced on a single Illumina MiSeq run, yielding an average of 71,000 paired-end reads (150 bp) per library corresponding to a median 800 x coverage per genome. After demultiplexing, reads from each library were trimmed for low quality stretches using sickle (v1.290, available at github.com/najoshi/sickle), then error-corrected and assembled with SPAdes (v2.5.0) using BayesHammer error correction and multiple k-mers (21,33,55,77) (Bankevich et al. 2012). Assembled scaffolds contained in each case one high-coverage scaffold matching existing amplicon, other scaffold in assembly corresponded to very low coverage partial sequences. To check for potential mis-assemblies, reads were mapped back to each mitochondrial genome using Bowtie2 (v2.2.1) (Langmead and Salzberg 2012) and inspected visually for coverage discrepancies (supplementary Data set S2, Supplementary Material online). Obtained sequences were also compared with cox1 and cox2 fragments to verify no sample mixing happened.

Annotation of Mitochondrial Genomes
Predicted ORFs were annotated and their delimitation refined by comparison with alignments of existing chaetognath mitochondrial sequences. Refined amino acid and nucleotide sequences were aligned back to each genome for annotation and determination of gene order (fig. 4 and supplementary fig. S7, Supplementary Material online).

Molecular Evolution and Population Genetic Analysis
Population genetic parameters were calculated using the EggLib python library (De Mita and Siol 2012). NeighborNet Networks were built using SplitsTree with K2P distances (Huson and Bryant 2006). Maximum Likelihood (ML) phylogenetic inferences were performed using RAxML assuming the GTR+Γ model for nucleotide data sets and the MtZoa+Γ model for amino acid data sets (Rota-Stabelli et al. 2009; Stamatakis 2014) (fig. 2A). Nonsynonymous and synonymous substitution rates were estimated using the yN00, site-model and branch model using the PAML package (supplementary table S4, Supplementary Material online) (Yang 2007).

Survey of Mitochondrial Genetic Diversity in Metazoa
NCBI popset database was searched for data sets corresponding to cox1 gene fragments (keyword: Cox1, COI, CO1), including a single species and at least ten sequences. We obtained 1581 records corresponding to 1,079 species. We focused on the 437 data sets associated with a referenced publication (pmid). A codon alignment was built using nucleotide translation discarding partial sequences and those containing degenerate bases (e.g., “N”, “Y”, or “R”). We computed nucleotide diversity statistics using the egg-lib library (De Mita and Siol 2012) and we selected the data set showing the highest diversity in each species (supplementary table S2, Supplementary Material online). The 50 data sets with highest levels of genetic diversity were manually curated for both the sequence alignments and corresponding papers.

Transcriptome Data
RNA was extracted from ~100 mixed embryonic stages and juveniles of S. cephaloptera using the Qiagen RNA micro kit (Invitrogen). Normalized cDNA libraries were built by BioS&T Inc. and sequenced according to manufacturer’s instruction on a GS GLX instrument (Roche).

Results
Extreme Mitochondrial Diversity in Two Chaetognath Morphospecies
We characterized mitochondrial variation of individual chaetognaths from three species sampled at single geographic localities in Sormiou (France) and Gullmar fjord (Sweden) (supplementary fig. S1, Supplementary Material online). All chaetognaths were examined under a microscope while still alive and identified to species level based on morphological characters. We sequenced mitochondrial gene fragments in more than 180 individuals, and full mitochondrial genomes in 54 representative individuals (table 1 and supplementary table S1, Supplementary Material online). We calculated nucleotide diversity (θ), which summarizes the average number of
nucleotide difference per site in a population sample, as well as uncorrected sequence divergence for the cox1 gene. In S. cephaloptera (N = 25), we found π = 0.14 and median sequence divergence of 17.99% (pairwise maximum = 23.68%). In S. elegans (N = 107), we found π = 0.18 and median sequence divergence of 20.80% (pairwise maximum = 24.96%). In both species, phylogenetic analysis of individual mitochondrial genes, such as 16S rRNA or cox1, revealed that this diversity is distributed across 8–11 highly divergent lineages (fig. 1; supplementary fig. S2 and Data set S1, Supplementary Material online). However, we did not detect such high levels of mitochondrial diversity in a third species, S. setosa (cox1: N = 54, π = 0.009, 0.91% median sequence divergence), which is a closely related species to S. elegans, and was sampled at the same location in Gullmar fjord (table 1).

We never recovered multiple products in cross-amplification experiments, nor consistent double Sanger chromatogram peaks, which would be indicative of multiple mtDNA copies per individual (Breton et al. 2007) or nuclear pseudogenes (Bensasson et al. 2001). All mitochondrial genome amplicons included the 13 genes typical of chaetognaths and we also did not find any stop codons or frameshift mutations despite global nucleotide diversities of 0.29 and 0.21 in S. cephaloptera and S. elegans, respectively (table 1). To further ascertain the validity of our PCR-based sequence data, we examined a transcriptome data set obtained from >100 pooled individuals of S. cephaloptera, all sampled from the same location as the genotyped individuals. We found transcripts of multiple mitochondrial genes, including cox1, with unambiguous assignment of all five lineages for which a whole mitogenome was sequenced in S. cephaloptera (supplementary Data set S2, Supplementary Material online). These multiple lines of evidence reject explanations such as PCR artifacts, the presence of multiple mitochondrial copies (heteroplasmy) or the presence of nuclear pseudogenes as possible explanations for the extreme levels of mitochondrial diversity observed here.

**Nuclear Variation Is Incongruent with Deep Mitochondrial Lineages**

We compared the patterns of mitochondrial variation with those inferred from three nuclear markers: the variable ribosomal protein L36a intron (fig. 1) and the more conserved 18S and 28S rRNA genes (table 1). The ribosomal genes 18S and 28S are commonly used to resolve interspecific relationships within the phylum because of their strong interspecific divergence (Gasmi et al. 2014; Papillon et al. 2006; Telford and Holland 1997). If deep mitochondrial lineages represent ancient, reproductively isolated, species, we would expect to see congruent divergence patterns across independently evolving nuclear loci (Coyne and Orr 2004; Hudson and Turelli 2003).

For S. elegans, we characterized variation at three nuclear loci: 18S and 28S markers showed no variation at all, but the nuclear intron of L36a was quite variable with 68 alleles detected in 37 individuals (table 1). A gene tree of the L36a intron shows a complete mixing of individuals with highly divergent mitochondrial genomes (fig. 1) and hence, displays incongruent patterns of divergence with the mitochondrial gene tree for this species. We also found significant levels of recombination in the nuclear intron data set, as would be expected for an interbreeding population of individuals (phi-test, P value = 0). Conversely, for S. cephaloptera we obtained information for only two nuclear markers and the results are inconclusive. The nuclear intron of L36a was sequenced for 30 individuals and we recovered 36 different alleles. These nuclear alleles arrange in four supported clades, which appear incongruent with deep mitochondrial lineages (fig. 1). We only sequenced eight individuals for 18S, which was quite variable in this species with 27 variable sites, but more genotyping data would be required to determine whether individuals belonging to deep mitochondrial clades in S. cephaloptera are diverged at nuclear loci. Nevertheless, phylogenetic trees combining our 18S and 28S sequences from genotyped individuals with those of other available chaetognath species (supplementary figs. S3 and S4, Supplementary Material online, respectively) show that divergences within morphospecies S. elegans and, to a lesser extent, also S. cephaloptera, are generally much lower than those observed between chaetognath species.

Inconsistency between mitochondrial and nuclear gene trees can be the result of hybridization between, usually closely related, species, resulting in introgression of mtDNA from one species to the other (Toews and Brelsford 2012). However, we consider this unlikely because in a gene tree based on clustered cox1 sequences from 29 chaetognath species (five species of which show deep mitochondrial clades >10% divergent), we found that all mitochondrial lineages in each species share a common ancestor (i.e., represent a monophyletic group) and hence are never shared between species (supplementary fig. S5, Supplementary Material online).

Among the three sampled nuclear markers with varying evolutionary rates, none showed patterns of divergence congruent with the observed deep mitochondrial lineages in either of two chaetognath species. More nuclear loci should be characterized though to fully assess whether completely or partially isolated gene pools are present in natural populations of S. cephaloptera and S. elegans.

**Extreme Divergence between Mitochondrial Lineages**

Mitochondrial genes annotated in sequenced genomes consistently show high divergence and congruent topologies as expected for a haploid, nonrecombining genome (supplementary Data set S2, Supplementary Material online). We thus examined phylogenetic relationships and divergence...
using the complete set of mitochondrial protein-coding sequences along with existing mitogenomic data from other chaetognath species (fig. 2A and supplementary fig. S6, Supplementary Material online) (Helfenbein 2004; Miyamoto et al. 2010a; Papillon et al. 2004). We found that this tree recovers with good support the accepted relationships between chaetognath species (Gasmi et al. 2014; Papillon et al. 2006), and unambiguously supports the monophyly of mitogenomes from all three species analysed in this study.

To put mitochondrial divergence levels into perspective, we compared the phylogenetic distance between chaetognath mitochondrial lineages with those inferred between the main vertebrate groups (fig. 2B). We found that maximum-likelihood distances between mitochondrial lineages of *S. cephaloptera* and *S. elegans* compare with those recovered within amniote and mammalian lineages, which diverged 312 and 170 Myr ago, respectively (Hedges and Kumar 2009). For instance, the distance between the Sor3 lineage to the common ancestor of *S. cephaloptera* is 0.707 substitution per site.

**Fig. 1.**—Incongruence of mitochondrial and nuclear lineages in chaetognaths. *Spadella cephaloptera* and *Sagitta elegans* sampled at single geographic sites (supplementary fig. S1, Supplementary Material online). Neighbor-net phylogenetic networks of mitochondrial 16S (left) and nuclear L36a intron (right) sequences reconstructed using K2P genetic distances. Numbering of individuals is the same for the two markers in each species and labels are colored according to mitochondrial lineage assignment. For nuclear sequences, a and b denote two different alleles recovered by cloning from the same individual. Scale represents expected nucleotide changes per site.
whereas the distance from human to the amniote ancestor is 0.671 substitutions per site. Similarly, the distance between human and platypus is 0.630 and that between lineage A and E in *S. elegans* is 0.717 substitutions per site. Although speculative, this analysis suggests that deep mitochondrial lineages in chaetognaths are associated with ancient divergence, strong mutation rate acceleration, or a combination of the two.
Mitochondrial Diversity Values Represent Extremes Amongst Animals

Nucleotide diversity estimates for two chaetognath species represent the highest values reported so far for any metazoan as established by surveying public databases for all available population cox1 data sets (fig. 3 and supplementary table S2, Supplementary Material online). We classified the data sets as single species or possible cryptic species by analysing the corresponding papers. S. elegans shows the highest level of synonymous diversity ($\pi_S = 0.646$) followed by S. cephaloptera ($\pi_S = 0.476$). Conversely, S. cephaloptera is the more variable species at the coding level ($\pi_N = 0.023$) followed by S. elegans ($\pi_N = 0.018$). A few data sets that were considered to result from cryptic speciation showed slightly higher diversity levels (open circles in fig. 3). We find that both chaetognath species harbor more intraspecific variation than several established cases of extreme mitochondrial diversity driven by (micro-)allopatric divergence, such as the crustacean Tigriopus californicus ($\pi_S = 0.404$) and the gastropod Cepaea nemoralis ($\pi_S = 0.418$) (Burton and Lee 1994; Thomaz et al. 1996).

Mitochondrial Rearrangements between Lineages

During the annotation of the mitochondrial genomes of Sagitta elegans, we observed gene order differences between individual genomes, which appear strictly associated with lineages. This suggests that beyond fast molecular divergence, these lineages underwent unusual structural rearrangements (fig. 4). We confirmed the validity of these structural changes by checking the read alignments. We found no coverage discontinuity associated with breakpoints and we recovered the same arrangements when sequencing multiple genomes from the same individuals (supplementary Data set S2, Supplementary Material online).

In S. elegans, at least two lineages exhibit rearrangements compared with the standard gene order involving the genes nad1, nad2, and cox3 in one case, and nad4L in the other (fig. 4B). Structural changes in mitogenomes also affect the size and proportion of intergenic regions, which are highly variable between lineages, ranging from 8.3% in lineage H to 25.4% in lineage G. Conversely, the fraction of noncoding DNA is remarkably stable within lineages with at most 3% variation (supplementary fig. S7, Supplementary Material online). These intergenic regions do not contain palindromic motifs which would be indicative of a repetitive element origin (Lavrov 2010).

Changes in gene order are commonly observed between species (Miyamoto et al. 2010a; Papillon et al. 2004), but to our knowledge such rearrangements have never been described as part of the diversity of a single animal species before. We also identified partial cox1 extra-copies inserted in intergenic regions of S. cephaloptera mitogenomes Sce-2 and Sce-6, which had accumulated coding substitutions (grey regions in fig. 4A). These duplicates support the model that considers gene duplications as intermediate steps in structural rearrangements of the mitochondrion (Moritz et al. 1987). Such dynamic rearrangements could be responsible for the extreme reduction in size and gene number of chaetognath mitogenomes compared with other bilaterian animals, with the loss of atp6, atp8, and tRNA genes (Helfenbein 2004; Papillon et al. 2004).

Patterns of Selection in Mitochondrial Genomes

Because high levels of coding nucleotide diversity could be the result of positive selection, we examined the patterns of nonsynonymous versus synonymous variation in mitochondrial genes (Oliveria et al. 2008). We found low or moderate $\pi_N/\pi_S$ ratios (average 0.170 for S. elegans and 0.049 for S. cephaloptera, fig. 5A and supplementary table S3, Supplementary Material online) as well as low intraspecific $d_N/d_S$ ratios between lineages (average 0.086 for S. elegans and 0.097 for S. cephaloptera, supplementary table S3, Supplementary Material online). We further tested whether individual lineages contribute unequally to the global $d_N/d_S$ estimates by assigning independent $d_N/d_S$ ratios to each lineage in individual genes. No particular lineage or gene showed any evidence of relaxation of selective pressure, though some genes seem more prone to higher coding variation than others in certain lineages, such as nad6 in Lineage G or nad3 in Lineage C (fig. 5B). Similarly, likelihood-ratio tests did not provide significant support for positive selection affecting specific lineages or sites (supplementary table S4,

---

**Fig. 3.**—Mitochondrial diversity values in chaetognaths are highest among animals. The nucleotide diversity was computed for 437 cox1 data sets extracted from the NCBI popsets database that are associated with a referenced publication. Data sets were classified as single species or possible cryptic species (closed and open circles, respectively).
Supplementary Material online). These analyses consistently indicate that positive selection is not responsible for the divergence of mitochondrial lineages in *S. cephaloptera* and *S. elegans*, instead, the low d_0/d_S values suggest that chaetognath mitochondrial lineages evolved mainly under the influence of purifying selection.

**Discussion**

We report the presence of highly divergent mitochondrial lineages in individuals of the chaetognaths *S. cephaloptera* and *S. elegans* sampled at single geographic sites. We further demonstrate that these deep mitochondrial lineages are split by molecular divergences equivalent to those observed among tetrapods (fig. 2) and contain structural rearrangements (fig. 4). The mechanism through which such divergent lineages could appear within these natural populations remains to be ascertained. Mainly, there are two possible interpretations of our results: deep lineages represent cryptic species and are reproductively isolated, or, alternatively, deep lineages are present within interbreeding populations.

Although cryptic speciation is a common explanation for such extreme mitochondrial divergence, we did not observe a pattern of divergence at any of the examined nuclear loci consistent with reproductive isolation (fig. 1; supplementary figs. S3 and S4, Supplementary Material online). Moreover, the high number of observed mitochondrial lineages in both species (more than eight) in sympatry would require a complicated scenario of multiple isolation events followed by secondary contact. Our present data set, however, does not incorporate a sufficient number of nuclear loci to decide between these evolutionary scenarios, and particularly, to confidently rule out the possibility that lineages are reproductively isolated. In addition, our sampling was limited to single geographic localities for both species, which makes it difficult to estimate which part of the molecular diversity may be attributable to spatial variation. Hence, extending our geographical sampling and acquiring a broader sampling of nuclear markers across the genome—for instance using RNA-seq or target capture methods (Gayral et al. 2013)—would be necessary to differentiate between the two alternative interpretations of our results.
If the alternative interpretation is correct and highly divergent lineages are present within interbreeding populations of chaetognaths, they likely appeared because of an accelerated mitochondrial mutation rate, and were maintained because of large population sizes. By considering synonymous and nonsynonymous rates of substitutions, we ruled out that these lineages emerged through positive selection. Instead, low or moderate \( \text{d}_\text{S}/\text{d}_\text{N} \) ratios were suggestive of the influence of strong purifying selection typical of species with large population sizes (supplementary table S4, Supplementary Material online) (Bazin et al. 2006; Meiklejohn et al. 2007; Ellegren and Galter 2016). Other cases of deeply divergent sympatric mitochondrial lineages without evidence of nuclear differentiation have been reported, for example, for the annelid \textit{Lumbricus rubellus} (five lineages up to 16% divergent) (Giska et al. 2015) and the bird \textit{Phoenicus phoenicus} (two lineages up to 5% divergent) (Hogner et al. 2012). In both cases, the authors concluded that the mitochondrial diversity was present within single highly polymorphic species rather than a complex of several cryptic species, and was unlikely the result of introgression. The presence of highly divergent mitochondrial lineages would fit theoretical expectations predicting that a large population size allows for the maintenance of ancestral polymorphisms (Hahn 2008; Lewontin 1974), and would challenge previous reports suggesting that the extent of mitochondrial variation across animals is limited (Bazin et al. 2006).

Since \textit{S. cephaloptera} and \textit{S. elegans} are distantly related in the chaetognath phylogenetic tree (belonging to the orders Phragmophora and Aphythomorpha, respectively (Papillon et al. 2006; Gasmi et al. 2014)) and have very distinct ecologies (benthic and planktonic, respectively), extreme mitochondrial diversity is probably not related to the particular demographic characteristics of species. Previous reports of deep mitochondrial lineages in other chaetognath species are consistent with this claim, even though these have often been interpreted as evidence of cryptic speciation (Kulagin et al. 2014; Miyamoto et al. 2010b). Not all chaetognath morphospecies harbor extreme levels of mitochondrial diversity though, as we show for the species \textit{S. setosa} (table 1). The reduced mitochondrial variation in this species (global \( \pi = 0.0088 \)) compared with other chaetognaths has been related to the fragmented, coastal populations of \textit{S. setosa}, which probably suffered severe population bottlenecks during Pleistocene glacial periods (Peijnenburg et al. 2004, 2005, 2006). As previously stated, more extensive geographic sampling for \textit{S. cephaloptera} and \textit{S. elegans} and the genotyping of a broader collection of nuclear markers could shed further light on the origin and maintenance of these deep mitochondrial lineages in chaetognaths.

An accelerated mitochondrial mutation rate could be related to the reduction of chaetognath mitogenomes and their propensity to structural rearrangements (Miyamoto et al. 2010a). Such a combination of extreme size reduction and accelerated mutation rate was also reported for a ctenophore (Pett et al. 2011). The size reduction and peculiar architecture of animal mitochondrial genomes compared with other eukaryotes, such as green plants, was attributed to the combined effect of genetic drift and accelerated mutation rates (Lynch et al. 2006). Hence, the processes that gave rise to the extreme mitochondrial diversity in chaetognaths could be similar to those governing mitochondrial evolution in other metazoans, but only differ by the extent of mutation rate acceleration and (effective) population sizes.

If highly divergent lineages are present within interbreeding populations, this would imply that cyto-nuclear interactions at the respiratory complexes are much less constrained than originally thought (Kenyon and Moraes 1997). Indeed, the interactions between mitochondrial and nuclear subunits should be robust enough to cope with multiple divergent mitochondrial lineages present in interbreeding populations.
mitochondrial genomes. Mitochondrial transplantation experiments originally demonstrated that cyto-nuclear interactions could be maintained between species, but were dramatically altered when divergence increased (e.g., to human and orang-utan), resulting in a decreased efficiency of respiratory processes (Kenyon and Moraes 1997). Similar effects were reported in hybrids of divergent Tigriopus californicus populations, which showed reduced respiratory fitness and altered gene expression (Ellison and Burton 2008a, 2008b). We hypothesize that chaetognaths cope with high levels of mitochondrial variation through an increased robustness in the interaction with nuclear subunits of oxidative phosphorylation complexes. How this increased robustness relates to the original genomic traits of chaetognaths, such as gene duplication, represents an interesting future topic of investigation.

We report the presence of highly divergent mitochondrial genomes in different individuals from the same morphospecies sampled at single geographic localities. Whether these highly divergent genomes are present within interbreeding populations or are attributable to a complex reproductive isolation scenario remains to be determined using a larger collection of nuclear loci and broader geographic sampling. Nevertheless, this finding challenges established views of the amount of mitochondrial diversity in animal species and populations. An assessment of intraspecific mitochondrial variation through an increased robustness in the interaction with nuclear subunits of oxidative phosphorylation complexes. How this increased robustness relates to the original genomic traits of chaetognaths, such as gene duplication, represents an interesting future topic of investigation.

We report the presence of highly divergent mitochondrial genomes in different individuals from the same morphospecies sampled at single geographic localities. Whether these highly divergent genomes are present within interbreeding populations or are attributable to a complex reproductive isolation scenario remains to be determined using a larger collection of nuclear loci and broader geographic sampling. Nevertheless, this finding challenges established views of the amount of mitochondrial diversity in animal species and populations. An assessment of intraspecific mitochondrial variation through an increased robustness in the interaction with nuclear subunits of oxidative phosphorylation complexes. How this increased robustness relates to the original genomic traits of chaetognaths, such as gene duplication, represents an interesting future topic of investigation.

Data Accessibility

- Assembled mitochondrial genomes have been deposited in Genbank under the accessions KP899748-KP899801 (details in supplementary table S1, Supplementary Material online).
- Sequences of genotyped markers are available under the accessions: KP843748-KP843841 and KP857119-KP857568 (details in supplementary table S1, Supplementary Material online).
- Transcriptome data was deposited with the accession PRJNA357934.
- Genotyped markers, alignments, and reconstructed trees are also available as supplementary Data sets S1 and S2, Supplementary Material online at https://figshare.com/s/a4706e9ad1ecea55880.

Supplementary Material

Supplementary data are available at Genome Biology and Evolution online.

Author Contribution

F.M., Y.L.P., and K.P. conceived the project and carried out the animal sampling. F.M. and S.L. performed the experiments. F.M., S.L., and Y.L.P. analyzed the sequencing data. F.M. and K.P. wrote the paper.

Acknowledgments

This research was supported by NWO-VENI grant 863.08.024 to K.P. and by ERC (FP7-2007–2013)/ERC grant [268513]11 funding F.M. The Wellcome Trust Centre for Human Genetics (WTCHG) is funded by Wellcome Trust grant 090532/Z/09/Z and MRC Hub grant G0900747.10/07. We thank L. Friis Møller and C. Hörnlein for help with sampling in Gullmar fjord in Sweden; E. Goetze, M. Taylor, I. Maeso, J. Paps, W. Renema, M. Schilthuizen, M. Malinsky, and P. W. H. Holland for commenting on the manuscript; E.J. Bosch for help with illustrations; P. Kuperus and C. Ribout for technical assistance and the WTCHG for the generation of the sequencing data.

Literature Cited


Associate editor: Dennis Lavrov