OIST OKINAWA INSTITUTE OF SCIENCE AND TECHNOLOGY GRADUATE UNIVERSITY
沖縄科学技術大学院大学

Extreme Mitogenomic Variation in Natural Populations of Chaetognaths

Extreme Mitogenomic Variation in Natural Populations of Chaetognaths

Ferdinand Marlétaz[1,2,*], Yannick Le Parco[3], Shenglin Liu[4], and Katja TCA Peijnenburg[4,5,*]

[1]Department of Zoology, University of Oxford, United Kingdom

[2]Molecular Genetics Unit, Okinawa Institute of Science and Technology, Onna, Japan

[3]Institut Méditerranéen d'Océanologie (CNRS UMR 7294), Aix-Marseille Université, Campus de Luminy, Marseille, France

[4]Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, The Netherlands

[5]Naturalis Biodiversity Center, Marine Biodiversity, Leiden, The Netherlands

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

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 2008b). 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 (Marlétaz 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 of copepods (Bone et al. 1991). Unambiguous fossils of chaetognaths have been described from the lower Cambrian (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 (Papillon et al. 2006; Gasmi et al. 2014). Transcriptomic studies revealed that chaetognath genomes underwent extensive gene duplication; many genes are transcribed in operons, often associated with trans-splicing (Marlétaz et al. 2008). The nuclear genome of chaetognaths also appears highly polymorphic, with widespread structural variation driven by mobile elements (Marlétaz et al. 2010). All chaetognaths are protandrous hermaphrodites with internal fertilization occurring while eggs are in the ovary. Though self-fertilization has been observed in the laboratory, cross-fertilization is considered to be the rule in nature (Alvinaro 1965; Pearre 1991; Peijnenburg et al. 2006) .

Previous studies examining the genetics of chaetognaths reported high levels of diversity and significant population structuring, but also revealed unusual patterns of mitochondrial evolution (Kulagin et al. 2014; Marlétaz et al. 2008, 2010; Miyamoto 2012; Miyamoto et al. 2010b; Peijnenburg et al. 2004, 2005, 2006). In this study, we focus on single populations of a benthic (*Spadella cephaloptera*) and two planktonic (*Sagitta elegans* and *Sagitta setosa*) chaetognaths from which we sampled a large number of invividuals (table 1). We show that extreme levels of mitogenomic variation can exist within natural populations without obvious evolutionary consequences.

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 (Lysekil, Sweden). All chaetognaths were identified under a microscope while still alive by taxonomic experts (FM and KTCAP). Only mature individuals without 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 (Finnzymes). 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 the *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*

Species/Location	Gene	n	K	π	θ_w	D_T
Spadella cephaloptera	*cox1*	25	25	0.1385	0.1223	0.5319
Calanque of	*16S*	30	29	0.1181	0.1149	0.1066
Sormiou, France	All genes[a]	5	5	0.2939	0.2470	1.4414
	18S[a]	8	8	0.0095	0.0102	−0.3362
	L36a intron[a]	30	36	0.0411	0.0443	−0.2605
Sagitta elegans	*cox1*	107	96	0.1775	0.0854	3.6069
Gullmar fjord,	*cox2*	108	101	0.2043	0.1113	2.7856
Sweden	All genes[a]	37	37	0.2081	0.1304	2.1922
	18S[a]	24	1	0.0000	0.0000	NA
	28S[a]	27	1	0.0000	0.0000	NA
	L36a intron[a]	37	68	0.0275	0.0590	−1.8220
Sagitta setosa	*cox1*	54	49	0.0088	0.0214	−2.0446
Gullmar fjord,	*cox2*	54	53	0.0138	0.0305	−1.9015
Sweden	All genes[a]	12	12	0.0088	0.0139	−1.6561
	18S[a]	20	2	0.0001	0.0002	−1.1643
	28S[a]	20	6	0.0006	0.0015	−1.7800

cox1, Cytochrome Oxidase 1; all genes, all coding mitochondrial genes; n, number of individuals; K, number of haplotypes/alleles; π, nucleotide diversity; θ_w, Watterson estimator; D_T, Tajima's D. For "all genes," average values are shown. For more details, see supplementary table S1, Supplementary Material online.

[a]Nonrandom sampling as selected individuals for divergent mitochondrial lineages were used. Hence, these estimates should not be considered as representative of population genetic diversity.

built 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× 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 expecting amplicon, other scaffolds 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). Neighbor-Net Networks were built using Splitstree4 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 Metazoans

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 (π), 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 $\pi = 0.14$ and median sequence divergence of 17.99% (pairwise maximum = 23.68%). In *S. elegans* ($N = 107$), we found $\pi = 0.18$ 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*: $N = 54$, $\pi = 0.009$, 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 mitochondrial 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. cephalptera* 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
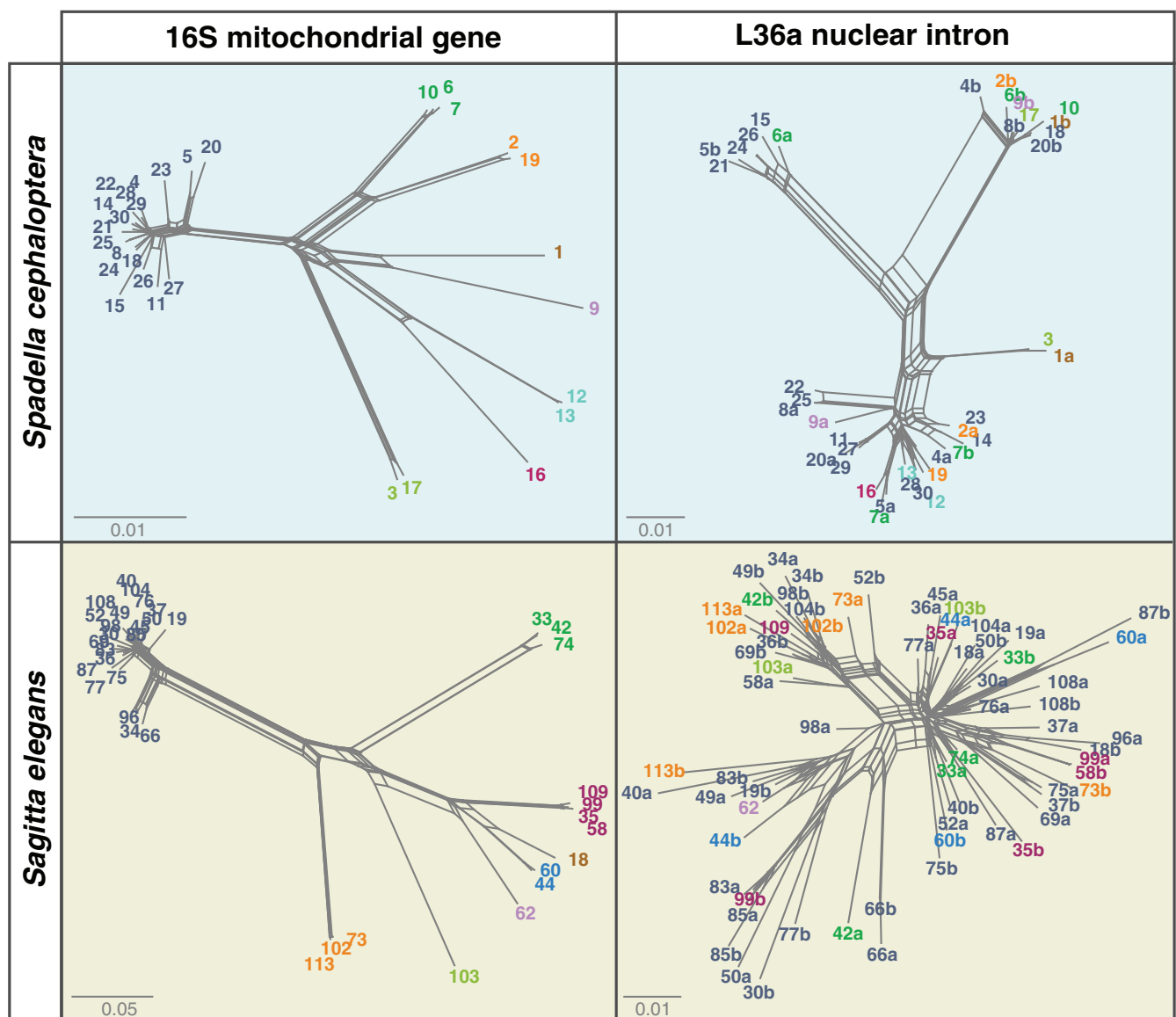
Fig. 1.—Incongruence of mitochondrial and nuclear lineages in chaetognaths *Spadella cephaloptera* and *Sagitta elegans* sampled at single geographic sites (supplementary fig. S1, Supplementary Material online). Neighbor-net phylogenetic networks of mitochondrial 16S (left) and nuclear L36a intron (right) 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.

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,
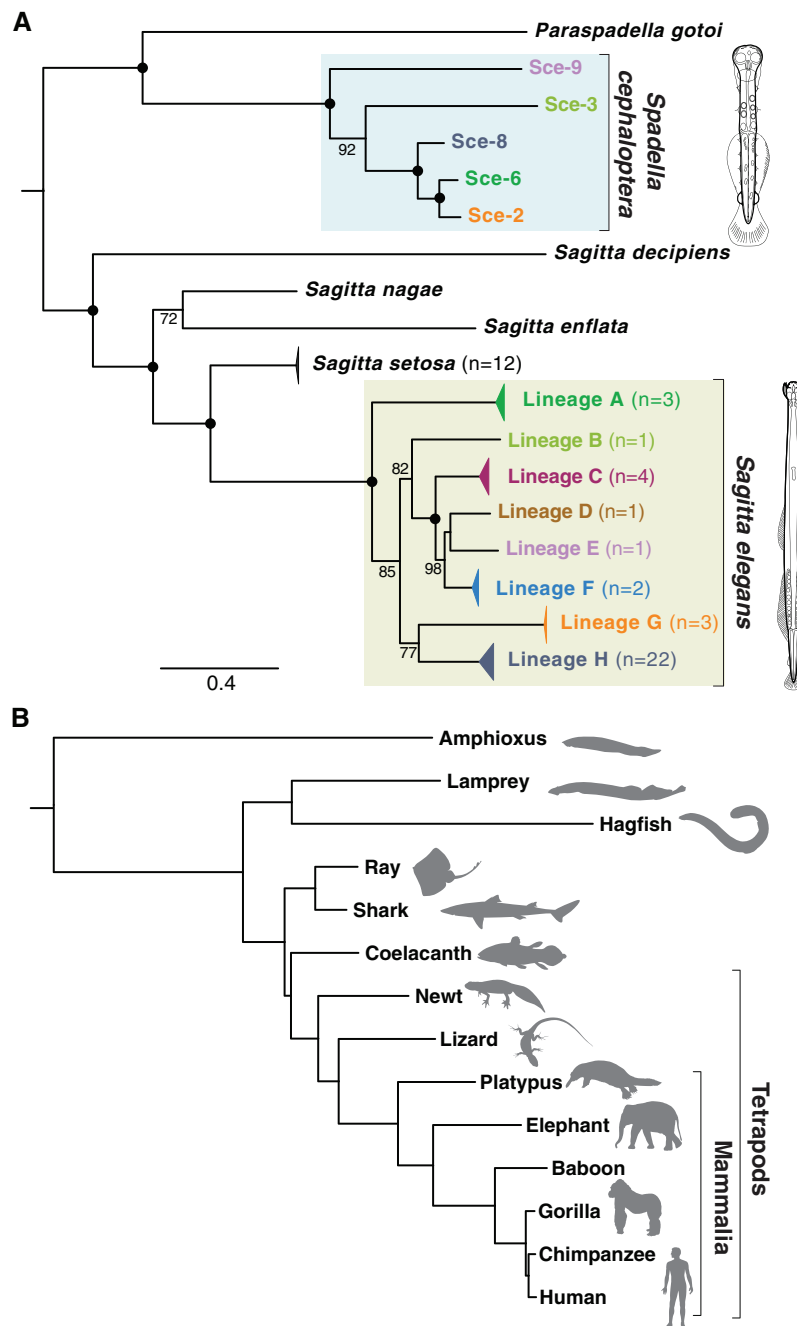
FIG. 2.—Mitochondrial divergence in chaetognaths and chordates. Phylogenetic trees based on the concatenation of all protein-coding mitochondrial genes in chaetognath lineages (A) and chordates (B) at the same scale (expected amino acid changes per site). Reconstructions were performed using Maximum Likelihood (MtZOA + Γ model). Maximum bootstrap support values are indicated by plain circles on nodes. In chordates, ML branch lengths were inferred from the alignment according to accepted topology. Sample sizes of *Sagitta elegans* lineages are indicated in brackets (detailed in supplementary fig. S6, Supplementary Material online).

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.

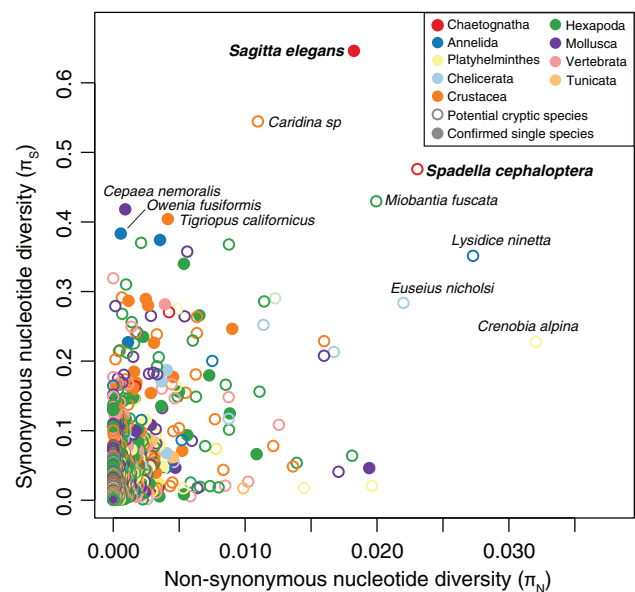## Mitochondrial Diversity Values Represent Extremes Amongst Animals

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


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).

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 (Oliveira et al. 2008). We found low or moderate π_N/π_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,
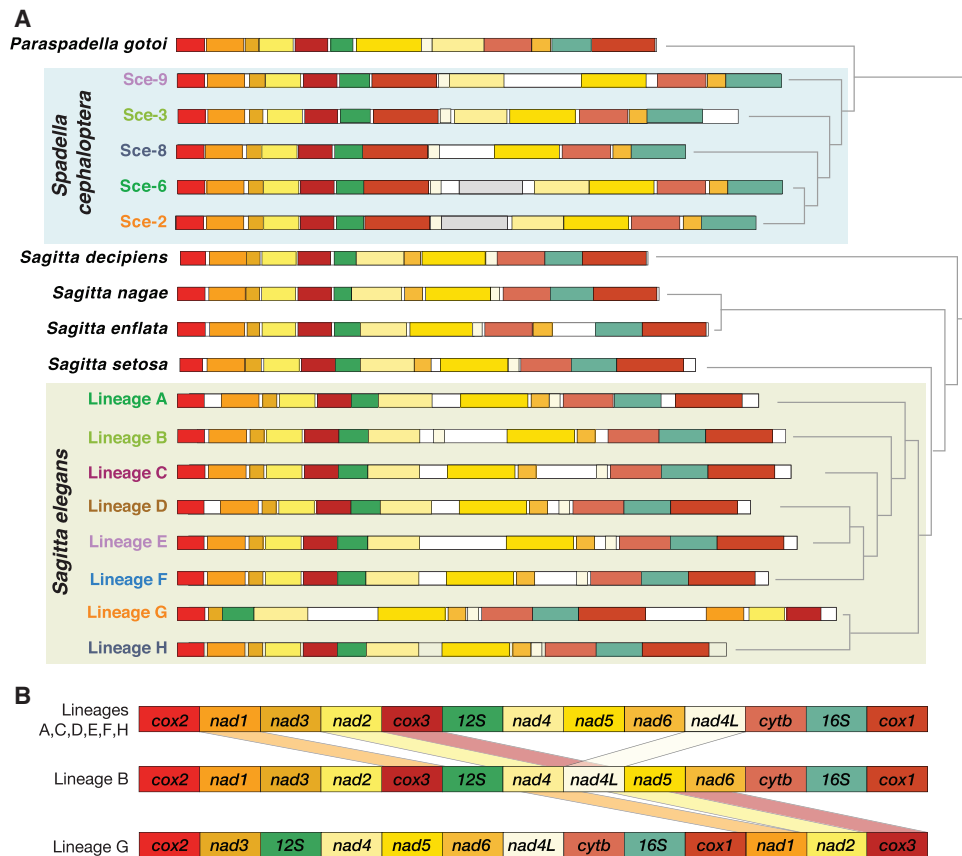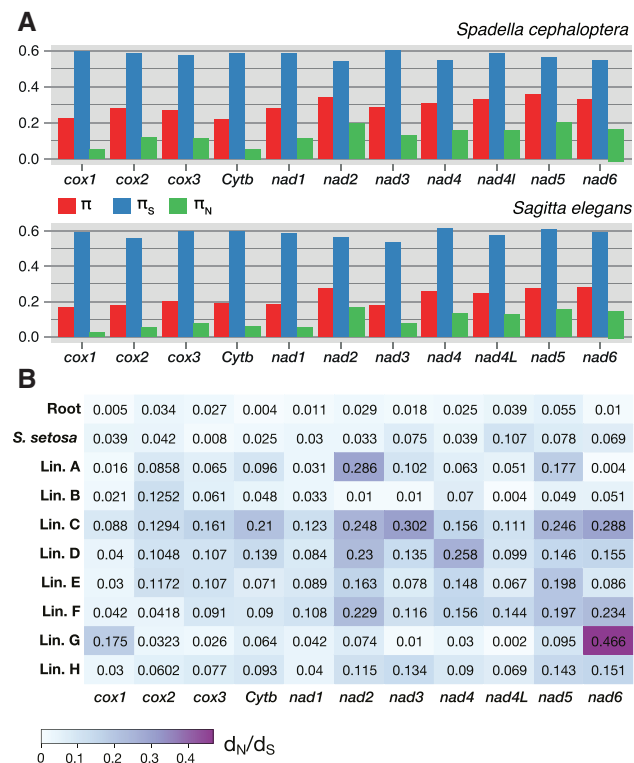
Fig. 4.—Structural variation between individual mitochondrial genomes of chaetognaths. (*A*) Annotated mitochondrial genomes with gene positions drawn to scale. The proportion of intergenic regions (white) varies from 8% to 26.4% in *Sagitta elegans* lineages and from 3.7% to 18.4% in *Spadella cephaloptera* lineages. *cox1* duplicates in *S. cephaloptera* are in grey. (*B*) Schematic representation of inferred mitochondrial gene order rearrangements in two *Sagitta elegans* mitochondrial lineages (B and G) with respect to the most frequent gene order (lineages A, C, D, E, F, and H). See also supplementary fig. S7, Supplementary Material online.

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_N/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.

A

Spadella cephaloptera

Sagitta elegans

(Bar charts showing nucleotide diversity π (red), π_S (blue), π_N (green) for cox1, cox2, cox3, Cytb, nad1, nad2, nad3, nad4, nad4l, nad5, nad6)

B

	cox1	cox2	cox3	Cytb	nad1	nad2	nad3	nad4	nad4L	nad5	nad6
Root	0.005	0.034	0.027	0.004	0.011	0.029	0.018	0.025	0.039	0.055	0.01
S. setosa	0.039	0.042	0.008	0.025	0.03	0.033	0.075	0.039	0.107	0.078	0.069
Lin. A	0.016	0.0858	0.065	0.096	0.031	0.286	0.102	0.063	0.051	0.177	0.004
Lin. B	0.021	0.1252	0.061	0.048	0.033	0.01	0.01	0.07	0.004	0.049	0.051
Lin. C	0.088	0.1294	0.161	0.21	0.123	0.248	0.302	0.156	0.111	0.246	0.288
Lin. D	0.04	0.1048	0.107	0.139	0.084	0.23	0.135	0.258	0.099	0.146	0.155
Lin. E	0.03	0.1172	0.107	0.071	0.089	0.163	0.078	0.148	0.067	0.198	0.086
Lin. F	0.042	0.0418	0.091	0.09	0.108	0.229	0.116	0.156	0.144	0.197	0.234
Lin. G	0.175	0.0323	0.026	0.064	0.042	0.074	0.01	0.03	0.002	0.095	0.466
Lin. H	0.03	0.0602	0.077	0.093	0.04	0.115	0.134	0.09	0.069	0.143	0.151

0   0.1   0.2   0.3   0.4 d_N/d_S

Fig. 5.—Nucleotide variation and molecular signatures of selection in mitochondrial genes. (*A*) Nucleotide diversities for all (π), noncoding (π_S), and coding (π_N) sites in 11 protein coding mitochondrial genes of *Spadella cephaloptera* (*N* = 5) and *Sagitta elegans* (*N* = 37) whole mitochondrial genomes. (*B*) d_N/d_S estimates per lineage using PAML "branch-model" in each mitochondrial gene of *Sagitta elegans*. Distinct d_N/d_S ratios were assigned for each gene to the internal branches of the tree ("base") leading to *Sagitta setosa* and to the multiple *Sagitta elegans* deep lineages (see also supplementary table S4, Supplementary Material online). 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.

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, the low or moderate d_N/d_S ratios are 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 Galtier 2016). Other cases of deeply divergent sympatric mitochondrial lineages without evidence of nuclear differentiation have been reported, for example, for the annelid *Lumbricus rubellus* (five lineages up to 16% divergent) (Giska et al. 2015) and the bird *Phoenicurus phoenicurus* (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 *S. cephaloptera* and *S. elegans* are distantly related in the chaetognath phylogenetic tree (belonging to the orders Phragmophora and Aphragmorphora, 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 *S. setosa* (table 1). The reduced mitochondrial variation in this species (global π = 0.0088) compared with other chaetognaths has been related to the fragmented, coastal populations of *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 *S. cephaloptera* and *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

of chaetognaths. 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 diversity in other animal groups, particularly those with dynamic mitochondrial genomes, such as ctenophores, urochordates, or brachiopods, may uncover other extreme cases of intraspecific variation (Pett et al. 2011). Until more of these hyperdiverse taxa are identified, chaetognaths represent an interesting and emerging model to understand the molecular evolution of animal mitochondrial genomes.

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/a4706e9adc1ecae55880.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Author Contribution

Acknowledgments

Literature Cited

Alvinaro A. 1965. Chaetognaths. Oceanogr Mar Biol Ann Rev. 3:115–194.

Bankevich A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19:455–477.

Bazin E, Glémin S, Galtier N. 2006. Population size does not influence mitochondrial genetic diversity in animals. Science 312:570–572.

Bensasson D, Zhang DX, Hartl DL, Hewitt GM. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol Evol. 16:314–321.

Bickford D, et al. 2007. Cryptic species as a window on diversity and conservation. Trends Ecol Evol. 22:148–155.

Bone Q, Kapp H, Pierrot-Bults AC. 1991. The biology of chaetognaths. Oxford: Oxford University Press.

Breton S, Beaupré HD, Stewart DT, Hoeh WR, Blier PU. 2007. The unusual system of doubly uniparental inheritance of mtDNA: isn't one enough? Trends Genet. 23:465–474.

Bucklin A, Steinke D, Blanco-Bercial L. 2011. DNA barcoding of marine metazoa. Ann Rev Mar Sci. 3:471–508.

Burton RS, Lee BN. 1994. Nuclear and mitochondrial gene genealogies and allozyme polymorphism across a major phylogeographic break in the copepod *Tigriopus californicus*. Proc Natl Acad Sci U S A. 91:5197–5201.

Coyne JA, Orr HA. 2004. Speciation. Sinauer Associates Incorporated.

De Mita S, Siol M. 2012. EggLib: processing, analysis and simulation tools for population genetics and genomics. BMC Genet. 13:27.

Ellegren H, Galtier N. 2016. Determinants of genetic diversity. Nat Rev Genet. 17:422–433.

Ellison CK, Burton RS. 2008a. Genotype-dependent variation of mitochondrial transcriptional profiles in interpopulation hybrids. Proc Natl Acad Sci U S A. 105:15831–15836.

Ellison CK, Burton RS. 2008b. Interpopulation hybrid breakdown maps to the mitochondrial genome. Evolution 62:631–638.

Galtier N, Nabholz B, Glémin S, Hurst GDD. 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Mol Ecol. 1–10.

Gasmi S, et al. 2014. Evolutionary history of Chaetognatha inferred from molecular and morphological data: a case study for body plan simplification. Front Zool. 11:84.

Gayral P, et al. 2013. Reference-free population genomics from next-generation transcriptome data and the vertebrate–invertebrate gap. 9:e1003457.

GIGA Community of Scientists, et al. 2014. The Global Invertebrate Genomics Alliance (GIGA): developing community resources to study diverse invertebrate genomes. J Hered. 105:1–18.

Gillespie JH. 2001. Is the population size of a species relevant to its evolution? Evolution 55:2161–2169.

Giska I, Sechi P, Babik W. 2015. Deeply divergent sympatric mitochondrial lineages of the earthworm Lumbricus rubellus are not reproductively isolated. BMC Evol Biol. 15:217.

Hahn MW. 2008. Toward a selection theory of molecular evolution. Evolution 62:255–265.

Hedges SB, Kumar S. 2009. The Timetree of Life. Oxford: Oxford University Press.

Helfenbein KG, Fourcade HM, Vanjani RG, Boore JL. 2004. The mitochondrial genome of Paraspadella gotoi is highly reduced and reveals that chaetognaths are a sister group to protostomes. Proc Natl Acad Sci USA. 101:10639–43.

Hogner S, et al. 2012. Deep sympatric mitochondrial divergence without reproductive isolation in the common redstart Phoenicurus phoenicurus. Ecol Evol. 2:2974–2988.

Hudson RR, Turelli M. 2003. Stochasticity overrules the 'three-times rule': genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. Evolution 57:182–190.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 23:254–267.

Kenyon L, Moraes CT. 1997. Expanding the functional human mitochondrial DNA database by the establishment of primate xenomitochondrial cybrids. Proc Natl Acad Sci U S A. 94:9131–9135.

Knoll AH, Carroll SB. 1999. Early animal evolution: emerging views from comparative biology and geology. Science 284:2129–2137.

Kulagin DN, et al. 2014. Spatial genetic heterogeneity of the cosmopolitan chaetognath Eukrohnia hamata (Mobius, 1875) revealed by mitochondrial DNA. Hydrobiologia 721:197–207.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359.

Lavrov DV. 2010. Rapid proliferation of repetitive palindromic elements in mtDNA of the endemic Baikalian sponge Lubomirskia baicalensis. Mol Biol Evol. 27:757–760.

Leffler EM, et al. 2012. Revisiting an old riddle: what determines genetic diversity levels within species? PLoS Biol. 10:e1001388.

Lewontin RC. 1974. The genetic basis of evolutionary change. Oxford: Columbia University Press.

Lynch M, Koskella B, Schaack S. 2006. Mutation pressure and the evolution of organelle genomic architecture. Science 311:1727–1730.

Marlétaz F, et al. 2006. Chaetognath phylogenomics: a protostome with deuterostome-like development. Curr Biol. 16:R577–R578.

Marlétaz F, et al. 2008. Chaetognath transcriptome reveals ancestral and unique features among bilaterians. Genome Biol. 9:R94.

Marlétaz F, Gyapay G, Le Parco Y. 2010. High level of structural polymorphism driven by mobile elements in the Hox genomic region of the Chaetognath Spadella cephaloptera. Genome Biol Evol. 2:665–677.

Matus DQ, et al. 2006. Broad taxon and gene sampling indicate that chaetognaths are protostomes. Curr Biol. 16:R575–R576.

Meiklejohn CD, Montooth KL, Rand DM. 2007. Positive and negative selection on the mitochondrial genome. Trends Genet. 23:259–263.

Miyamoto H. 2012. Global phylogeography of the deep-sea pelagic chaetognath Eukrohnia hamata. Prog Oceanogr. 104:99–109.

Miyamoto H, Machida RJ, Nishida S. 2010a. Complete mitochondrial genome sequences of the three pelagic chaetognaths Sagitta nagae, Sagitta decipiens and Sagitta enflata. Comp Biochem Physiol 5:65–72.

Miyamoto H, Machida RJ, Nishida S. 2010b. Genetic diversity and cryptic speciation of the deep sea chaetognath Caecosagitta macrocephala (Fowler, 1904). Deep-Sea Res. 57:2211–2219.

Moritz C, Dowling T, Brown W. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annu Rev Ecol Syst. 18:269–292.

Oliveira DCSG, Raychoudhury R, Lavrov DV, Werren JH. 2008. Rapidly evolving mitochondrial genome and directional selection in mitochondrial genes in the parasitic wasp Nasonia (hymenoptera: pteromalidae). Mol Biol Evol. 25:2167–2180.

Papillon D, Perez Y, Caubit X, Le Parco Y. 2004. Identification of chaetognaths as protostomes is supported by the analysis of their mitochondrial genome. Mol Biol Evol. 21:2122–2129.

Papillon D, Perez Y, Caubit X, Le Parco Y. 2006. Systematics of Chaetognatha under the light of molecular data, using duplicated ribosomal 18S DNA sequences. Mol Phyl Evol. 38:621–634.

Pearre S. 1991. Growth and reproduction. In: The biology of chaetognaths. Oxford: Oxford University Press. 61–75.

Peijnenburg KTCA, Breeuwer JAJ, Pierrot-Bults AC, Menken SBJ. 2004. Phylogeography of the planktonic chaetognath Sagitta setosa reveals isolation in European seas. Evolution 58:1472–1487.

Peijnenburg KTCA, Fauvelot C, Breeuwer JAJ, Menken SBJ. 2006. Spatial and temporal genetic structure of the planktonic Sagitta setosa (Chaetognatha) in European seas as revealed by mitochondrial and nuclear DNA markers. Mol Ecol. 15:3319–3338.

Peijnenburg KTCA, Goetze E. 2013. High evolutionary potential of marine zooplankton. Ecol Evol. 3:2765–2781.

Peijnenburg KTCA, Haastrecht EK, Fauvelot C. 2005. Present-day genetic composition suggests contrasting demographic histories of two dominant chaetognaths of the North-East Atlantic, Sagitta elegans and S. setosa. Mar Biol. 147:1279–1289.

Pett W, et al. 2011. Extreme mitochondrial evolution in the ctenophore Mnemiopsis leidyi: Insight from mtDNA and the nuclear genome. Mitochondrial DNA 22:130–142.

Piganeau G, Eyre-Walker A. 2009. Evidence for variation in the effective population size of animal mitochondrial DNA. PLoS One 4:e4396.

Reed DH, Frankham R. 2003. Correlation between fitness and genetic diversity. Conserv Biol. 17:230–237.

Rota-Stabelli O, Yang Z, Telford MJ. 2009. MtZoa: a general mitochondrial amino acid substitutions model for animal evolutionary studies. Mol Phyl Evol. 52:268–272.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

Telford MJ. 2004. Evolution: affinity for arrow worms. Nature 431:254–256.

Telford MJ, Holland PW. 1997. Evolution of 28S ribosomal DNA in chaetognaths: duplicate genes and molecular phylogeny. J Mol Evol. 44:135–144.

Thomaz D, Guiller A, Clarke B. 1996. Extreme divergence of mitochondrial DNA within species of pulmonate land snails. Proc Biol Sci. 263:363–368.

Toews DPL, Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. Mol Ecol. 21:3907–3930.

Vannier J, Steiner M, Renvoisé E, Hu S-X, Casanova J-P. 2007. Early Cambrian origin of modern food webs: evidence from predator arrow worms. Proc Biol Sci. 274:627–633.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.

Associate editor: Dennis Lavrov