Transcription, Signaling Receptor Activity, Oxidative Phosphorylation, and Fatty Acid Metabolism Mediate the Presence of Closely Related Species in Distinct Intertidal and Cold-Seep Habitats

Jelle Van Campenhout1,2,*, Ann Vanreusel1, Steven Van Belleghem3,4, and Sofie Derycke1,4

1Research Group Marine Biology, Biology Department, Ghent University, Belgium
2Department of Biology, Center for Molecular Phylogenetics and Evolution (CeMoFe), Ghent University, Biology Department, Belgium
3Terrestrial Ecology Unit, Biology Department, Ghent University, Belgium
4OD Taxonomy and Phylogeny, Royal Belgian Institute of Natural Sciences, Brussels, Belgium
*Corresponding author: E-mail: jelle.ugent@gmail.com.

Accepted: November 28, 2015

Data deposition: The raw sequencing reads were deposited at the Short Read Archive of the National Center for Biotechnology Information (NCBI) under bioproject SRP050898 containing biosamples SRS748702 (Run SRR1657929) and SRS748703 (Run SRR1657930).

Abstract

Bathyal cold seeps are isolated extreme deep-sea environments characterized by low species diversity while biomass can be high. The Håkon Mosby mud volcano (Barents Sea, 1,280 m) is a rather stable chemosynthetic driven habitat characterized by prominent surface bacterial mats with high sulfide concentrations and low oxygen levels. Here, the nematode Halomonhystera hermesi thrives in high abundances (11,000 individuals 10 cm−2). Halomonhystera hermesi is a member of the intertidal Halomonhystera disjuncta species complex that includes five cryptic species (GD1-5). GD1-5’s common habitat is characterized by strong environmental fluctuations. Here, we compared the transcriptomes of H. hermesi and GD1, H. hermesi’s closest relative. Genes encoding proteins involved in oxidative phosphorylation are more strongly expressed in H. hermesi than in GD1, and many genes were only observed in H. hermesi while being completely absent in GD1. Both observations could in part be attributed to high sulfide concentrations and low oxygen levels. Additionally, fatty acid elongation was also prominent in H. hermesi confirming the importance of highly unsaturated fatty acids in this species. Significant higher amounts of transcription factors and genes involved in signaling receptor activity were observed in GD1 (many of which were completely absent in H. hermesi), allowing fast signaling and transcriptional reprogramming which can mediate survival in dynamic intertidal environments. GC content was approximately 8% higher in H. hermesi coding unigenes resulting in differential codon usage between both species and a higher proportion of amino acids with GC-rich codons in H. hermesi. In general our results showed that most pathways were active in both environments and that only three genes are under natural selection. This indicates that also plasticity should be taken in consideration in the evolutionary history of Halomonhystera species. Such plasticity, as well as possible preadaptation to low oxygen and high sulfide levels might have played an important role in the establishment of a cold-seep Halomonhystera population.

Key words: transcriptomics, adaptive radiation, phenotypic plasticity, evolutionary biology, cryptic species, nematodes.

Introduction

One of the most important goals in evolutionary biology is to comprehend how and why living organisms diversify (Pfenning et al. 2010). Dispersal allows organisms to colonize new suitable habitats and found new populations. Speciation in the new habitat has its origin in the establishment of reproductive barriers between source and founder population (Coyne and Orr 2004). While gene flow between populations is often considered as a homogenizing evolutionary force that can prevent conspecific populations from diverging (Walsh and Mena 2013), ecological selection can also initiate speciation in the presence of gene flow (Seehausen et al. 2014).

© The Author 2015. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
Therefore, restriction of gene flow and ecological differentiation between habitats are two drivers which can mediate speciation, thereby, paving the way for adaptive evolution of newly founded populations (Mayr 1963).

Adaptive evolution is the process in which natural selection will select for advantageous alleles that provide a higher fitness to the organism thereby causing changes in allele frequencies in a population. Adaptive evolution is influenced by mutation rate, recombination, genetic drift, selection, etc. (reviewed in Olson-Manning et al. 2012) each of which cause genetic patterns that are traceable in sequence data. In addition, individuals will respond to changes in their environment by remodeling their transcriptome, allowing individuals to quickly respond to environmental conditions. This is referred to as phenotypic plasticity where one genotype can produce multiple phenotypes. While some studies define phenotypic plasticity as nonheritable (Ellner et al. 2011), gene expression can also be regulated by epigenetic modifications which are heritable (Surani et al. 1993; Cubas et al. 1999; Chong and Whitelaw 2004; Bird 2007; Goldberg et al. 2007). In the case of cryptic species, that is, morphologically identical but genetically distinct species, the degree of plasticity in morphological traits is limited, despite the fact that they can be found in quite distinct habitats (Vrijenhoek 2009). Understanding the biology of cryptic species in different environments requires information on genetic adaptive traits and the degree of plasticity in other traits than morphology (Gittenberger and Gittenberger 2011).

Cryptic diversity is common in marine environments, which may be attributed to the use of chemical cues for mate recognition (Stanhope et al. 1992; Palumbi 1994; Lonsdale et al. 1998). The existence of cryptic species has been described in different orders of marine nematodes (Derycke et al. 2005, 2007, 2010), usually the most abundant and diverse Metazoa in marine benthic communities (Heip et al. 1985; Lambshead and Boucher 2003). In this work, we focus on the genus Halomonhystera Andrassy (2006) belonging to the order of the Monhysterida which is currently situated outside the five clades (I–V) proposed by Blaxter (2011) and placed in clade 5c based on the numerical system of van Megen et al. (2009). The genus has a widespread geographical distribution and has been identified from deep-sea environments (Van Gaever et al. 2006; Portnova et al. 2010) and intertidal regions (Trotter and Webster 1983; Vranken, Herman, et al. 1988; Mokievsky et al. 2005; Derycke et al. 2007). Nuclear and mitochondrial sequence data led to the discovery of five cryptic Halomonhystera disjuncta species (GD1-5) with only limited morphometric differences in the Western Scheldt estuary (Derycke et al. 2007; Fonseca et al. 2008). Halomonhystera was also reported as the dominant nematode genus in the sulfide-rich bacterial mats of the Nyegga pockmark (Nordic Norwegian margin, 730 m) (Portnova et al. 2010) and the Håkon Mosby mud volcano (HMMV, 1,280 m, Barent Sea slope) (Van Gaever et al. 2006). However, a recent study revealed subtle morphological and clear genetic differences between the HMMV and intertidal Halomonhystera species (Van Campenhout et al. 2013): uncorrected $p$-distances between HMMV and GD1 ranged between 19.1% and 25.2% for the cytochrome oxidase c subunit I gene part, whereas uncorrected $p$-distances of the nuclear 18S gene ranged between 2.4% and 2.9% and the nuclear ITS-D2D3 concatenated genes showed a divergence of 9.5% (Van Campenhout et al. 2013). The HMMV Halomonhystera has consequently been described as a new species, Halomonhystera hermesi (Tchesunov et al. 2015). Phylogenetic analysis revealed a close relationship between $H$. hermesi and both GD1 and GD4 (Van Campenhout et al. 2013). For Halomonhystera, a deep-sea invasion from intertidal regions has been hypothesized followed by adaptive evolution as a result of absence of gene flow between both environments (Van Campenhout et al. 2013) and strong selective forces in the deep-sea habitat. GD1 is hypothesized to share the most recent ancestor with $H$. hermesi as it appears to be more resistant to bathyal seep conditions compared with GD2-3 (Van Campenhout et al. 2014). In addition, previous studies have revealed a high tolerance of $H$. disjuncta to low food- and high heavy metal concentrations (Vranken et al. 1985, 1989; Vranken, Tire, et al. 1988). These studies clearly highlight that in addition to the genetic differences between GD1 and $H$. hermesi, the intertidal species also have the plasticity to survive deep-sea conditions. Halomonhystera hermesi thrives at the HMMV, a cold, stable chemosynthetically based deep-sea environment (Van Gaever et al. 2006). The microbial mats, the habitat of $H$. hermesi, are characterized by limited oxygen penetration (~1 mm) and high sulfide concentrations (up to 4 mM) (de Beer et al. 2006; Van Gaever et al. 2006). Even though episodic events such as extensive methane venting and mud flows occur, the upflow velocity in the bacterial mats is limited to 0.3–1.0 m yr$^{-1}$ (de Beer et al. 2006). In contrast, GD1 thrives on decaying algae in the Western Scheldt, an ephemeral habitat with daily fluctuations in temperature, light, salinity, and frequent inundation due to tidal activity. We, therefore, consider the HMMV habitat to be more stable compared with the intertidal habitat of GD1.

The detailed knowledge on the biology and phylogenetic relationships of these two species, together with their occurrence in well characterized and distinct habitats render them an excellent system to study the molecular mechanisms allowing species to thrive in different habitats. To this end, we compared the transcriptome between the intertidal (GD1) and deep-sea nematode ($H$. hermesi) related species. Evaluation of gene expression will contribute to our understanding of which pathways are important in both species with respect to their environment. Additionally, pinpointing genes that are under natural selection will contribute to our understanding of adaptive evolution of both species. Moreover, its intertidal occurrence subjects GD1 to strong physical, chemical, and biological gradients indicating that it can rapidly adjust to
environmental changes. Organisms in fluctuating environments must constantly sense and adapt to environmental changes and must be able to quickly respond to these changes. This may be achieved by transcriptional regulation (Kussell and Leibler 2005). In contrast, stable environments generally favor specialist species adapted to that specific niche (Miner 2005). In view of the clear differences in (abiotic) conditions in both habitats, we expected to find substantial differences in the molecular pathways activated in both habitats. More specifically, we expected to find one/more genes expressed and more active pathways in the intertidal than in the deep-sea habitat, and two/an overrepresentation of pathways involved in transcription and environmental information processing in the intertidal GD1 species.

**Materials and Methods**

**Sample Collection**

Deep-sea sediment samples were collected, using a TV-guided multicorer, from the active methane seeping Håkon Mosby mud volcano (HMMV; situated west of the Barents Sea (72.1°N 14.73°E) at an average water depth of 1,280 m) during the Maria S. Merian cruise 2010 (MSM 16/2). Core samples were immediately frozen in liquid nitrogen and stored at −80°C.

Intertidal monospecific GD1 nematodes reared in the lab were transferred to rehydrated (salinity of 25 practical salinity unit [psu]) deaerated macroalgae (*Fucus vesiculosus*) and allowed to grow for approximately 4 weeks in three litter bags (mesh size 200 μm) at 8°C and a salinity of 25 psu. Following this period, litter bags were positioned in the Paulina salt marsh (Western Scheldt, The Netherlands, at 51° 20.9’N, 3° 43.5’E) allowing nematodes to acclimatize to the environment for 72 h before retrieval. Macroalgae were washed over two stacked sieves (top sieve: 1 mm, bottom sieve: 32 μm) with sieved (32 μm) seawater and fauna was retained on the 32 μm sieve. Meiofauna from the sieve was then collected with sterile artificial seawater (25 psu) and replicate samples were immediately frozen in liquid nitrogen and stored at −80°C.

**Nematode Retrieval and RNA Isolation**

The frozen first deep-sea sediment slice (0–1 cm) of three locations (Station MSM16-2_802-1, 72° 0.17’N 14° 43.88’E; Station MSM16-2_829-1, 72° 0.16’N 14° 43.94’E; Station MSM16-2_831-1, 72° 0.14’N 14° 43.94’E) and all intertidal samples were washed separately over a 32 μm-mesh sieve. *Halomonhystera* nematodes were extracted by density centrifugation at 4°C and 3,000 × g, using Ludox (a colloidal silica polymer; specific gravity 1.18) as a flotation medium (Heip et al. 1985). β-mercaptoethanol, in a final concentration of 0.14 M, was added to avoid RNA breakdown. Nematodes were captured on a 32 μm sieve after centrifugation and washed again with a sterile formamide-β-mercaptoethanol (0.143 M) solution and finally retained in the same solution. Replicate samples of both deep-sea and intertidal samples contained approximately 1,500–2,000 nematodes. To minimize the effect of different life stages, only adult *H. hermesi* and GD1 nematodes of each replicate were morphologically identified and randomly and manually picked out using a binocular microscope in a 4°C climate room. Samples from GD1 or *H. hermesi* were pooled, prior to RNA extraction, into 50 μl RLT buffer (RNeasy Mini kit, Qiagen Inc., Düsseldorf, Germany) to which we added β-mercaptoethanol in a final concentration of 0.143 M. Each sample (deep sea and intertidal) contained approximately 5,000 nematodes and such pooling of replicates was a necessary practical consideration to obtain sufficient RNA.

The cuticle and cell membranes of nematodes and cells were disrupted by vortexing with beads (1 min at 5,000 rpm). Total RNA was extracted using the RNeasy Mini kit (Qiagen Inc., Düsseldorf, Germany) according to the manufacturer’s instructions. Genomic DNA was removed by on-column digest with DNase I. RNA was quantified by measuring the optical density (OD) at 260 nm using a NanoDrop spectrophotometer (2000 UV-Vis, Thermo Fisher Scientific, Inc.). RNA purity was assessed at an absorbance ratio of OD260/280 and OD260/230. RNA integrity was analyzed on a 1% agarose gel stained with ethidium bromide.

**cDNA Library Construction, Illumina Sequencing, and Quality Filtering**

Samples were sent to Bio S&T Inc. (Canada) to perform cDNA synthesis and normalization allowing us to sequence low expressed genes. Briefly, the cDNA library was constructed with about 8 μg of total RNA using a modified Clontech SMART cDNA Library Construction kit using oligo(dT) primers. Double stranded cDNAs were obtained by primer extension and purified. To increase the ability to sequence low expressed transcripts, cDNAs were normalized, reamplified, and purified (Bio S&T, Canada). Further library construction and sequencing were performed by Genomics Core, UZ Leuven, Belgium. Briefly, the cDNA was sheared using the Covaris M220 machine with a maximum fragment length of 800 bp and a fragment size peak around 450–500 bp. cDNA libraries were end repaired, and NEXTflex adaptors from BIOO Scientific were ligated using the SPRWorks robot (Beckman) followed by an AMPure bead purification to remove adaptor dimers. Libraries were amplified using a 12 cycle polymerase chain reaction step with NEXTflex primer mix. An additional AMPure bead purification was performed. The resulting cDNA libraries were pooled and paired-end sequenced on a single Illumina MiSeq 2*250 bp lane. Because of a low read diversity within deep-sea reads (troubling for the nucleotide incorporation imaging software), this cDNA library was mixed in with other samples.
and sequenced on the same Illumina MiSeq 2*250 bp machine. Data from both runs were pooled.

Low quality read ends (<Q20) were trimmed with FastX 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit, last accessed November 1, 2015). MiSeq adapters were trimmed only at the end (at least 10 bp overlap and 90% match) with cutadapt 1.0 (Martin 2011). Using FastX 0.0.13 and ShortRead 1.14.4 (Morgan et al. 2009), short reads (length <50 bp), poly A-reads (more than 90% of the bases equal A), ambiguous reads (containing N), low quality reads (more than 50% of the bases <Q25), and artifact reads (all but three bases in the read equal one base type) were removed in succession. Because of quality drop for all reverse reads after 150 bp, these reads were all cut after 150 bp. Finally, cDNA primers were removed using the CLC Genomics Workbench 6.0.3 trim sequences tool. The raw sequencing reads were deposited at the Short Read Archive of the National Center for Biotechnology Information (NCBI) under bioproject SRP050898 containing biosamples SRS748702 (Run SRR1657929) and SRS748703 (Run SRR1657930).

De Novo Assembly and BLAST

De novo assembly was conducted using the CLC Genomics Workbench 6.0.3 de novo assembly tool according to the manufacturer’s preset features. Contigs shorter than 200 bp were removed. To remove redundant transcripts and retain a set of contigs each representing a putatively unique isoform (unigenes), the open-source program CD-HIT-EST (Li and Godzik 2006) was used to cluster contigs at a 95% sequence identity. The largest contig of each cluster was retained in the data set. The coding sequences (CDS) were predicted using the open reading frame (ORF)-predictor server (Min et al. 2005) with a 200 bp CDS cutoff.

BLAST+ (Camacho et al. 2009) was used to subject the remaining unigenes to a similarity search against NCBI’s non-redundant (nr) database (standard genetic code) using the BLASTx algorithm (Altschul et al. 1990), with a cutoff e-value of $10^{-3}$ and a maximum of 20 BLAST hits. The top BLASTx hit was used to assign taxonomy to each unigene using a custom perl script (Get_classification.pl, Supplementary Material online). Unigenes classified as Nematoda were selected for further analysis.

Functional Annotation (KEGG) and Pathway Reconstruction

Prediction of pathways expressed in both data sets was done by using known metabolic and signaling pathways previously found in Eukaryotes and nematodes using the Kyoto Encyclopedia of Genes and Genomes (KEGG) automated Annotation Server (KAAS) (Moriya et al. 2007). Unigenes were submitted to KAAS with the preset Eukaryotes. *Caenorhabditis briggsae, Brugia malayi, Loa loa, and Trichinella spiralis* were added as additional sequence templates; the option “basis of single-best hit” was selected. KAAS annotates transcripts with KEGG orthology (KO) identifiers representing an orthologous group of genes linked to an object in the KEGG pathways and BRITE hierarchy (Mao et al. 2005; Moriya et al. 2007). Unigenes from both transcriptomes with the same KO identifier, that is, functional annotation, were considered to be shared between both species.

The KO identifiers were mapped against KEGG pathways to reconstruct pathways (http://www.genome.jp/kegg/mapper.html, last accessed April 30, 2015). KEGG pathways involved in human diseases were not taken into further consideration. KO identifiers were additionally mapped against BRITE hierarchical classifications of protein families.

To extract pathways that are over- and underrepresented in the deep-sea and intertidal habitats, respectively, KEGG pathways and protein families of unique unigenes from each transcriptome were determined separately using the Cytoscape (Shannon et al. 2003) plugin BINGO (Maere et al. 2005) with both transcriptomes combined as a reference set. P values were false-discovery rate (FDR)-corrected.

Functional Gene Ontology Annotation

To retrieve Gene Ontology (GO) terms describing biological processes, molecular functions, and cellular components (Ashburner et al. 2000), the publicly available BLAST2GO-platform (Conesa et al. 2005) was used. Annotations were conducted for unigenes with default parameters (e-value $\geq 10^{-6}$, Annotation cutoff $\geq 55$, and a GO weight of 5). BLAST2GO was also used to retrieve InterPro (conserved patterns in sequences) annotations and merged with GO terms for a wide functional range cover in annotation.

Because some unigenes without KO annotation displayed a GO annotation, the following step was taken to increase the amount of previously defined shared unigenes (based on KO identifier). Unigenes without KO identifier but with a GO annotation and a CDS of at least 200 bp from one transcriptome were aligned against both the shared unigenes and unigenes without KEGG annotation of the other transcriptome using the PROmer pipeline of the MUMmer 3.0 software (Kurtz et al. 2004) with default parameters. Unigenes with an in-frame PROmer hit, minimum similarity of 75% and an alignment length of minimum 300 bp were additionally considered as shared. Determining overrepresented GO-terms in unique unigenes was performed as described before. Similarly, we used BINGO to determine whether each complete GO data set was enriched in level 2 GO terms compared with both data sets combined.
Positive Selection, Amino Acid Composition, and Codon Usage

Orthologs present in both transcriptomes were determined by selecting unigenes with a CDS of minimal 200 bp which were then aligned and compared using the PROmer pipeline with default parameters. We only retained in-frame PROmer hits (similarity ≥ 75% and an alignment length of 300 bp) that matched the same KO identifier. The full length CDS of filtered orthologous gene pairs were then aligned using ClustalW (Thompson et al. 1994), with default parameters, provided in the Molecular Evolutionary Genetics Analysis version 6.0 (MEGA6) software package (Tamura et al. 2013) and uncorrected p-distances were calculated.

Positive selection was tested using the ratios of the rates of nonsynonymous substitutions per nonsynonymous sites (Ka) to the number of synonymous substitutions per synonymous site (Ks) using KaKs_Calculator 2.0 (Wang et al. 2010) with the γ-MYN model. This model incorporates previously unaddressed dynamic features of evolving DNA sequences in transition/transversion rate, nucleotide frequency, and unequal transitional substitution. The model further accounts for DNA sequences unequal substitution rates among different sites (Wang et al. 2009).

The alignment of orthologous gene pairs was also used to analyze GC content, amino acid (AA) composition, and codon usage. Amino acid frequencies were determined by summing the occurrence of each AA per translated CDS (tCDS) and dividing it through the total amount of AA present for each tCDS. Codon frequencies per AA were obtained by summing the occurrence of each single codon per tCDS and dividing it through the respective AA count of the tCDS. GC content, AA composition, and codon frequencies per gene were determined using a custom perl script (AA_counter.pl and Codon_counter.pl, Supplementary Material online) and averaged over all unigenes. To investigate whether significant differences were present between deep-sea and intertidal transcriptomes, mean values of both H. hermesi and GD1 corresponding orthologous unigenes were subtracted. The H. hermesi and GD1 data sets were then pooled and randomly permuted into two equal groups after which GC content, AA- and codon frequencies were calculated and subtracted from each other. This was repeated 100,000 times with custom R-scripts (GC_permutation.R, AA_permutation.R and Codon_permutation.R, Supplementary Material online) in R version 3.0.2 (R Core Team 2013) and the obtained values were seen as a null distribution for the GC content, codon, and AA frequencies. The P value is then calculated as the fraction of how many times the permuted difference is equal or more extreme than the observed difference and was FDR-corrected. If the observed difference fell outside the range of the null distribution, no P values could be calculated and P values were considered to be <E^{-6}.

Differential Gene Expression between Shared Unigenes

Deep-sea and intertidal contigs with a nematode top-BLAST hit were selected and used to generate a single new assembly using the de novo assembly tool in CLC Genomics Workbench 6.0.3. The assembly was treated in the same manner as described above (CD-HIT-EST clustering, BLASTx and functional annotation). Cleaned deep-sea and intertidal reads were mapped against the full assembly using the RNA-seq tool in CLC Genomics Workbench 6.0.3. (minimum length fraction = 0.9 and minimum similarity fraction = 0.8). Values for relative expression were based on mean RPKM (Reads Per Kilo-base of exon model per Million mapped reads) values. A statistical analysis on proportions (Kal’s test) was performed and P values were FDR-corrected. The RNA-seq analysis resulted in five data sets: 1) uniquely expressed genes in H. hermesi, 2) uniquely expressed genes in GD1, 3) shared unigenes similarly expressed in both species, 4) differentially overexpressed unigenes in H. hermesi, and 5) differentially overexpressed unigenes in GD1. Data set 1 and 2 were not analyzed further as this was already performed for the complete transcriptome. To increase the validity of the three other RNA-seq data sets, only unigenes with a KO identifier which was present in both original nematode transcriptomes and present in only one RNA-seq data set were retained. Over- and underrepresented KEGGs pathways, protein families, and GO annotations of differentially expressed genes were determined as described above.

Results

Sequencing Results

Illumina MiSeq 2*250 bp cDNA sequencing generated a total of 3.47 and 2.47 Gb for the deep-sea and intertidal samples respectively (supplementary table S1, Supplementary Material online). After quality assessment and data filtering, 4,605,081 and 3,105,224 pair-end reads remained for deep-sea and intertidal samples, respectively (supplementary table S1, Supplementary Material online). After quality assessment and data filtering, 4,605,081 and 3,105,224 pair-end reads remained for deep-sea and intertidal samples, respectively (supplementary table S1, Supplementary Material online). De novo assembly yielded a lower amount of deep-sea contigs (57,636) than intertidal contigs (72,678). In terms of unigenes, 54,767 deep-sea and 68,537 intertidal unigenes with a respective N50 of 933 and 720 bp were found. The size of the deep-sea unigenes ranged from 200 to 149,303 bp with a mean length of 750 whereas the intertidal unigenes ranged from 200 to 8,783 bp with a mean length of 618 bp (supplementary table S1, Supplementary Material online). De novo assembly yielded a lower amount of deep-sea contigs (57,636) than intertidal contigs (72,678). In terms of unigenes, 54,767 deep-sea and 68,537 intertidal unigenes with a respective N50 of 933 and 720 bp were found. The size of the deep-sea unigenes ranged from 200 to 149,303 bp with a mean length of 750 whereas the intertidal unigenes ranged from 200 to 8,783 bp with a mean length of 618 bp (supplementary table S1, Supplementary Material online).
**BLAST Results and Classification**

BLASTx results showed that 33,079 (60.40%) deep-sea and 42,409 (61.89%) intertidal unigenes had at least one BLAST result with an e-value $\leq 10^{-3}$. Surprisingly, a substantial portion of the unigenes did not give a nematode top-BLAST hit. Instead, most deep-sea unigenes classified as Bacteria (61.01%) whereas intertidal unigenes were mostly classified as Eukaryotes (92.52%) (fig. 1A). Of the intertidal unigenes that were classified as Eukaryota, only 49.51% could be assigned to the Metazoa (fig. 1B). The remaining part (40.31%) mainly belonged to Viridiplantae. In contrast, the deep-sea eukaryote unigenes were nearly all assigned to the Metazoa (93%; fig. 1B). Within the Metazoa, the composition of both data sets was highly similar: 67.99% deep-sea and 68.15% intertidal unigenes could be assigned to the Ecdysozoa, 16.02% deep-sea and 18.19% intertidal unigenes were assigned to the Chordata, and 8.95% deep-sea and 7.80% intertidal unigenes were designated as Lophotrochozoa (fig. 1C). The majority of the Ecdysozoa unigenes were classified as nematodes: 5,498 (79.95%) unigenes for the deep-sea and 8,588 (64.96%) for the intertidal data set (fig. 1D). Overall, 16.62% and 20.28% of the deep-sea and intertidal unigenes, respectively, were assigned to nematodes.

**Nematode Assembly**

Based on their nematode top-BLAST hit 5,498 deep-sea (*H. hermesi*) and 8,588 intertidal (GD1) unigenes were retained for further analysis. *Halomonhysterahermesi* unigenes had a maximum length of 7,235 bp and an average length of 943.07 bp. This was comparable to GD1 unigenes which had...
Closely Related Species in Distinct Intertidal and Cold-Seep Habitats

Table 1
Summary of BLAST Results, Assembly Statistics, ORF Prediction, and Annotation of Unigenes with a Nematode Top BLAST Hit of Two Targeted Locations and Species: HMMV (deep sea, Halomonhystera hermesi) and Western Scheldt (intertidal, GD1)

<table>
<thead>
<tr>
<th>Stage</th>
<th>Deep Sea</th>
<th>Intertidal</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLAST result</td>
<td>Unigenes with at least one BLAST result</td>
<td>33,079</td>
<td>42,409</td>
</tr>
<tr>
<td>Nematode assembly</td>
<td>Contigs with nematode top BLAST hit (unigenes)</td>
<td>5,498</td>
<td>8,588</td>
</tr>
<tr>
<td>N50 length (bp)</td>
<td>1,378</td>
<td>1,169</td>
<td></td>
</tr>
<tr>
<td>Max length (bp)</td>
<td>7,253</td>
<td>8,783</td>
<td></td>
</tr>
<tr>
<td>Average length (bp)</td>
<td>943.07</td>
<td>837.39</td>
<td></td>
</tr>
<tr>
<td>Average coverage per base</td>
<td>14.07</td>
<td>20.43</td>
<td></td>
</tr>
<tr>
<td>GC content unigenes (%)</td>
<td>56.11%</td>
<td>48.71%</td>
<td></td>
</tr>
<tr>
<td>ORF prediction</td>
<td>Unigenes with predicted ORF (CDS ( \geq 200 \text{bp} ))</td>
<td>5,227</td>
<td>8,102</td>
</tr>
<tr>
<td>KEGG annotation</td>
<td>Unigenes assigned to KO identifier</td>
<td>2,985</td>
<td>4,548</td>
</tr>
<tr>
<td>Unique KO identifiers</td>
<td>1,648</td>
<td>1,246</td>
<td>2,894</td>
</tr>
<tr>
<td>Go annotation</td>
<td>Mapped to GO terms</td>
<td>5,014</td>
<td>7,822</td>
</tr>
<tr>
<td>Unigenes annotated with GO terms</td>
<td>3,657</td>
<td>5,636</td>
<td>9,293</td>
</tr>
</tbody>
</table>

a maximum length of 8,783bp and an average length of 837.39bp (table 1). The amount of unigenes with a CDS longer than 200bp was lower for H. hermesi (5,227) than for GD1 (8,102). GC content of the H. hermesi unigenes was higher than that of the GD1 unigenes (56.11% vs. 48.71%, respectively).

Functional Annotation of the Full Transcriptomes

Using the KAAS, 2,985 H. hermesi unigenes (54.29%) were assigned to 1,648 KO identifiers, whereas 4,548 GD1 unigenes (52.96%) were assigned to 2,166 KO identifiers. Both transcriptomes shared 1,246 KO identifiers corresponding to 2,499 H. hermesi- and 3,244 GD1 unigenes. In addition, 402 KO identifiers were uniquely found in H. hermesi (corresponding to 487 unigenes), and 920 KO identifiers were uniquely found in GD1 (corresponding to 1,304 unigenes). These genes are further addressed as uniquely expressed unigenes. The KO identifiers were mapped to 245 shared pathways, and to three and ten unique pathways for H. hermesi and GD1, respectively (full data set in supplementary table S2, Supplementary Material online). However, only 1–3 unigenes were mapped to these unique pathways. The same three pathways were most represented in deep-sea and intertidal nematodes: “protein processing in endoplasmic reticulum,” “endocytosis,” and “spliceosome” (table 2). Of all shared pathways, 40 had a higher amount of unigenes mapped for H. hermesi than for GD1 (supplementary table S2, Supplementary Material online). "Oxidative phosphorylation" was the pathway with the highest differences in mapped proteins, that is, eight additional H. hermesi unigenes were mapped against this pathway compared with GD1. A higher amount of GD1 unigenes were mapped to 156 pathways compared with H. hermesi. The highest amount of additional GD1 unigenes (i.e., 23) mapped against the Neuroactive ligand-receptor interaction pathway. Equal amounts of GD1 and H. hermesi unigenes were mapped against 49 pathways (supplementary table S2, Supplementary Material online). At the level of protein families (full data in supplementary table S3, Supplementary Material online), “enzymes” was by far the most abundant category for both species (table 2).

Over/Underrepresentation of KEGG Pathways and Protein Families in Uniquely Annotated H. hermesi- and GD1 Unigenes

Unigenes that were uniquely found in H. hermesi or GD1 were considered to be characteristic of the deep-sea and intertidal environment, respectively, and were used to investigate pathways that were over/underrepresented in a particular habitat. Six KEGG pathways: "oxidative phosphorylation," “ubiquinone” and “other terpenoid-quione biosynthesis,” “benzoate degradation,” “terpenoid backbone biosynthesis,” “biosynthesis of ansamycins” and “fatty acid elongation” were significantly overrepresented in H. hermesi unique unigenes (P values < 0.05), all belonging to the metabolism category (fig. 2A). The unique unigenes were most enriched in the oxidative phosphorylation (P = 2.92E-3) belonging to the overrepresented energy metabolism class (P = 4.01E-3; fig. 2A). A closer inspection of our data revealed that the observed overrepresentation was due to NADH dehydrogenase (ubiquinone) 1 alpha/beta subcomplex subunits, several F- and V- type ATPase subunits, and succinate dehydrogenase, involved in the electron transport chain at the inner mitochondrial membrane. Three KEGG pathways were significantly underrepresented: “tight junction,” “circuitry entrainment,” and “salivary secretion.” For GD1, over- and underrepresented pathways in which unique unigenes were involved belonged to a wider spectrum of molecular pathways compared with the deep-sea transcriptome (fig. 2B). The most significant overrepresented pathway in GD1 uniquely annotated unigenes belonged to the neuroactive...
ligand-receptor interaction pathway ($p = 2.24 \times 10^{-22}$; fig. 2B) and was mostly the result of G-protein coupled receptors in the GD1 transcriptome which were not expressed in *H. hermesi*. Six other overrepresented pathways were “tropane,” “piperidine and pyridine alkaloid biosynthesis,” “signaling pathways regulating pluripotency of stem cells, ribosome,” “sulfur relay system,” “ubiquitin mediated proteolysis,” and “mismatch repair.” Many more pathways were underrepresented in GD1 compared with underrepresented *H. hermesi* pathways (supplementary table S4, Supplementary Material online). The top 3 were “fatty acid metabolism” ($P = 1.07 \times 10^{-4}$), “tight junctions” ($P = 3.54 \times 10^{-4}$), and “fatty acid biosynthesis” ($P = 7.01 \times 10^{-4}$).

Only protein families involved in genetic information processing were significantly ($P = 4.42 \times 10^{-5}$) overrepresented in *H. hermesi* unique genes as a result of overrepresented unique unigenes involved in “transcription factors,” “proteasome,” “t-RNA biogenesis,” and “ribosome biogenesis” (fig. 3A), Table 2

**Table 2**
Top 20 Selected KEGG Metabolic Pathways of Putative Proteins Mapped Using KEGG Database

<table>
<thead>
<tr>
<th>KEGG pathways</th>
<th>Full transcriptome</th>
<th>Unique unigenes</th>
<th>RNAseq unigenes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>H. hermesi</em></td>
<td><em>GD1</em></td>
<td><em>H. hermesi</em></td>
</tr>
<tr>
<td>Protein processing in endoplasmic reticulum</td>
<td>52</td>
<td>60</td>
<td>11</td>
</tr>
<tr>
<td>Endocytosis</td>
<td>46</td>
<td>60</td>
<td>10</td>
</tr>
<tr>
<td>Spliceosome</td>
<td>44</td>
<td>54</td>
<td>9</td>
</tr>
<tr>
<td>Oxidative phosphorylation</td>
<td>43</td>
<td>35</td>
<td>17</td>
</tr>
<tr>
<td>Rap1 signaling pathway</td>
<td>41</td>
<td>52</td>
<td>7</td>
</tr>
<tr>
<td>Purine metabolism</td>
<td>38</td>
<td>53</td>
<td>8</td>
</tr>
<tr>
<td>Focal adhesion</td>
<td>38</td>
<td>41</td>
<td>7</td>
</tr>
<tr>
<td>MAPK signaling pathway</td>
<td>37</td>
<td>45</td>
<td>11</td>
</tr>
<tr>
<td>Carbon metabolism</td>
<td>37</td>
<td>39</td>
<td>8</td>
</tr>
<tr>
<td>Lysosome</td>
<td>37</td>
<td>37</td>
<td>5</td>
</tr>
<tr>
<td>Regulation of actin cytoskeleton</td>
<td>36</td>
<td>46</td>
<td>5</td>
</tr>
<tr>
<td>PI3K-Akt signaling pathway</td>
<td>35</td>
<td>50</td>
<td>5</td>
</tr>
<tr>
<td>cAMP signaling pathway</td>
<td>35</td>
<td>43</td>
<td>3</td>
</tr>
<tr>
<td>Oxytocin signaling pathway</td>
<td>34</td>
<td>38</td>
<td>4</td>
</tr>
<tr>
<td>RNA transport</td>
<td>33</td>
<td>42</td>
<td>10</td>
</tr>
<tr>
<td>Ras signaling pathway</td>
<td>32</td>
<td>42</td>
<td>9</td>
</tr>
<tr>
<td>AMPK signaling pathway</td>
<td>32</td>
<td>33</td>
<td>7</td>
</tr>
<tr>
<td>cGMP—PKG signaling pathway</td>
<td>31</td>
<td>35</td>
<td>3</td>
</tr>
<tr>
<td>Insulin signaling pathway</td>
<td>29</td>
<td>41</td>
<td>5</td>
</tr>
<tr>
<td>Adrenergic signaling in cardiomyocytes</td>
<td>28</td>
<td>26</td>
<td>3</td>
</tr>
</tbody>
</table>

| Protein families                                   |                 |                |                |                |
|----------------------------------------------------|-----------------|-----------------|-----------------|
| Enzymes                                            | 708             | 901            | 161            | 354            | 59 | 44 | 14 |
| Exosome                                            | 160             | 188            | 22             | 50             | 12 | 21 | 9  |
| Transporters                                       | 123             | 144            | 22             | 43             | 7  | 6  | 0  |
| Protein kinases                                    | 111             | 139            | 19             | 47             | 10 | 12 | 3  |
| Chromosome                                         | 105             | 157            | 27             | 79             | 3  | 10 | 3  |
| Spliceosome                                        | 93              | 118            | 21             | 46             | 12 | 6  | 0  |
| Peptidases                                         | 91              | 111            | 20             | 40             | 8  | 6  | 1  |
| Protein phosphatases and associated proteins       | 75              | 95             | 17             | 37             | 4  | 4  | 0  |
| Mitochondrial biogenesis                           | 65              | 77             | 21             | 33             | 8  | 4  | 1  |
| Ubiquitin system                                   | 65              | 102            | 22             | 59             | 3  | 2  | 0  |
| Ribosome biogenesis                                | 59              | 78             | 16             | 35             | 11 | 10 | 0  |
| Transcription machinery                            | 57              | 74             | 10             | 27             | 2  | 5  | 2  |
| Chaperones and folding catalysts                   | 51              | 54             | 12             | 15             | 7  | 10 | 0  |
| Ion channels                                       | 49              | 61             | 14             | 26             | 2  | 0  | 1  |
| Transcription factors                              | 43              | 71             | 17             | 45             | 1  | 1  | 0  |
| Cytoskeleton proteins                              | 43              | 52             | 8              | 17             | 3  | 3  | 2  |
| DNA repair and recombination proteins              | 36              | 52             | 12             | 28             | 2  | 3  | 1  |
| Transfer RNA biogenesis                            | 34              | 46             | 11             | 23             | 1  | 3  | 1  |
| Proteasome                                         | 33              | 30             | 10             | 7              | 3  | 7  | 0  |
| Cell adhesion molecules and their ligands          | 32              | 48             | 7              | 23             | 1  | 2  | 0  |

**Note.**—Data of the full nematode transcriptome, the unique unigenes, equally (EE), and overexpressed (OE) RNAseq unigenes for *Halomonhystera hermesi* and GD1 is represented.
whereas exosome involved proteins were underrepresented. The two most overrepresented protein families in GD1 unique unigenes were G-protein-coupled receptors \( (P = 5.26 \times 10^{-16}) \) and transcription factors \( (P = 3.32 \times 10^{-9}; \text{fig. 3B}) \). Except for the overrepresentation of prenyltransferases, all other overrepresented protein families are involved in genetic information processing \( (P = 3.43 \times 10^{-11}) \).

**GO Classification of the Full Transcriptomes**

GO terms were assigned to 3,657 (72.94%) \( H. \) hermesi and 5,636 (72.05%) GD1 unigenes which could be classified into three categories: cellular component, molecular function, and biological process (fig. 4). The BINGO analysis of each GO data sets separate compared with both data sets combined did not reveal any overrepresented level 2 GO-terms. As such, GO annotation were highly similar between intertidal and deep-sea samples and reflected very well annotation patterns observed in the nematode *Caenorhabditis elegans* (depicted from geneontology.org). GO-annotation yielded 6,108 \( H. \) hermesi- and 9,591 GD1 GO cellular component annotations, 5,086 \( H. \) hermesi and 7,417 GD1 molecular function annotations, and 15,981 \( H. \) hermesi and 23,692 GD1 GO annotations for biological process. Within the cellular component category 32.95% and 32.69% of GD1 and \( H. \) hermesi unigenes were assigned to “Cell,” followed by “Organelle” (22.48% and 21.45%), “Membrane” (19.78% and 20.95%) and “Macromolecular Complex” (13.43% and 12.45%) (fig. 4). GO term assignment of \( H. \) hermesi- and GD1 unigenes, respectively, within the molecular function category, was dominated by “Binding” (40.59% and 40.61%) and “Macromolecular Complex” (13.43% and 12.45%) (fig. 4). “Receptor activity” and “nucleic acid binding transcription factor activity” were significantly overrepresented in GD1 unique unigenes \( (P \text{ values} = 3.18 \times 10^{-5} \text{ and } 1.80 \times 10^{-6}) \). Finally, within the biological process category 12 GO-terms captured more than 90% of the assignments.

**Fig. 2.**—Visualization of overrepresented KEGG pathways in unique \( H. \) hermesi (A) and GD1 (B) unigenes. Each hexagon represents either a KEGG pathway or higher order category according to BRITE hierarchy. White hexagons are not significant categories while colored ones represent the FDR \( P \) values.
of which “Single-Organism Process,” “Cellular Process,” and “Metabolic Process” were the three largest classes for both transcriptomes (fig. 4).

Overrepresentation of GO Terms in Unique Unigenes

The addition of selected PROmer hits resulted into 2,492 H. hermesi- and 4,752 GD1 unique unigenes. *Halomonhysterahermesi* unique unigenes only showed overrepresented GO terms from the molecular function category (P values in table 3) related to “enzyme inhibitor- and peptidase regulator activity” (supplementary fig. S1, Supplementary Material online). Much more overrepresented GO terms in GD1 unique unigenes were found (supplementary table S5, Supplementary Material online). Several GO terms involved in transcription, for example, transcription factor complex (cellular component, supplementary fig. S2, Supplementary Material online), sequence-specific DNA binding (molecular function, supplementary fig. S3, Supplementary Material online), and regulation of transcription (biological process, supplementary fig. S4, Supplementary Material online), were overrepresented in GD1 (P values in table 3). Moreover, overrepresented GO terms dealing with signaling pathways, for example, receptor- and kinase activity were prominent in both the molecular function and biological process category for GD1.

Positive Selection, Amino Acid Composition, and Codon Usage

Filtering PROmer hits based on similarity (≥75%) and alignment length (≥300 bp) resulted into 1,053 hits with the same functional annotation in both data sets. GC content of *H. hermesi* unigenes (58.97%) was significantly

---

**Fig. 3.**—Overrepresented protein families (BRITE mapping) in unique *H. hermesi* (A) and GD1 (B) unigenes. Parent-child relations are represented by arrows between hexagons. Hexagons' colors represent FDR P values from white, being not significant, to red, being highly significant.
higher \( P < E^{-6} \) compared with corresponding GD1 sequences (50.96%). Genetic divergence (uncorrected \( p \)-distances) between orthologous pairs ranged from 8.3% to 40.1%. Mean values of Ka, Ks and Ka/Ks ratio (fig. 5A) were 0.125, 3.266 and 0.0453, respectively. Of the orthologous pairs only three had a Ka/Ks which were mapped to three different KEGG pathways: “proteasome (20S proteasome subunit beta 3),” “Spliceosome (splicing factor U2AF 65 kDa subunit),” and “tight junctions (Myosin heavy chain).”

In both transcriptomes, GC rich codons were more frequently used, but this was especially true for the deep-sea transcriptome (fig. 5B). All codon differences were significant (supplementary table S6, Supplementary Material online). The relative frequency of amino acids (AAs) with GC-rich codons (Alanine, and Arginine) was significantly higher (both \( P \) values = 0.00105) in \( H. hermesi \) (fig. 5B). In contrast, GD1 had significantly higher relative frequency of AAs with AU-rich codons (fig. 5C) such as Isoleucine \( (P = 0.00105) \), Lysine \( (P = 0.00672) \), and Asparagine \( (P < E^{-6}) \). In addition, GD1 had higher frequencies of Threonine \( (P = 0.01467) \), Histidine \( (P = 0.0117) \), and Leucine \( (P = 0.0129) \) were higher in \( H. hermesi \). Summary statistics of the null distribution and \( P \) values can be found in supplementary table S7, Supplementary Material online.

**Differential Expression of Shared Genes between Intertidal and Deep-Sea Nematodes**

The de novo assembly, based on \( H. hermesi \)- and GD1 unigenes collectively, resulted in 3,345 unigenes with a CDS larger 200 bp. We identified 740 and 1,693 unigenes exclusive to \( H. hermesi \) and GD1, respectively, and 1,014 unigenes expressed in both species of which 845 were significantly differentially expressed (Kal’s Z-test). After applying the KO filtering on the 1,014 unigenes, we retained 142 and 118 unigenes that were differentially overexpressed in \( H. hermesi \) and GD1, respectively, whereas 90 were equally expressed.

We further focus on those genes that are involved in KEGG pathways and protein families that were shown to be over/underrepresented, allowing us to now compare gene expression levels of shared genes in these categories. Seven shared unigenes that are involved in transcription had a significantly higher gene expression for GD1 compared with four \( H. hermesi \).
hermesi overexpressed unigenes. Out of nine shared unigenes involved in oxidative phosphorylation, six revealed to have higher gene expression levels for H. hermesi (fig. 6). We observed very high gene expression differences for cytochrome c oxidase subunit I (COX1) between both species. Moreover, two subunits of the electron transport chain complex III, ubiquinol-cytochrome c reductase cytochrome c1 subunit (CYC1), and ubiquinol-cytochrome c reductase core subunit 2 (QCR2) had significantly higher gene expression values in H. hermesi.

**Discussion**

Halomonhystera hermesi and GD1 have a very similar morphology but were considered different species based on their genetic differentiation at mitochondrial and nuclear loci (Van Campenhout et al. 2013). The phylogenetic position of H. hermesi amongst the intertidal species suggests that H. hermesi has originated from shallow water relatives followed by adaptive evolution in the deep sea. However, no typical cold-seep adaptations such as ecto- and/or endosymbionts or sulfur crystal were observed (Van Gaever et al. 2006, 2009), which suggests that adaptations are situated at the molecular level.

Our results revealed that both species use the same molecular pathways to cope and survive in their distinct environment. Nevertheless, we found a higher amount of unigenes and active pathways in GD1, and most shared unigenes were differentially expressed between both species. Gene expression can be influenced by environmental factors (Lobo 2008) and our results revealed that different pathways were overrepresented in each environment suggesting that these pathways are of high importance to survive in the respective environment. The experimental procedure to collect specimens from either habitat may have added additional differences in gene expression between both habitats, and we may also have missed some very low expressed genes due to the normalization step. In addition, sex ratio differences between both populations might have influenced our results. Nevertheless, we believe that our approach is the most solid and feasible way for comparing the transcriptomes of both species in such different habitats.

**Transcription and Signaling Receptor Activity Are More Strongly Expressed in Intertidal Environments**

Our data disclosed a higher amount of unigenes, more overrepresented GO terms and a wider spectrum of molecular
**Fig. 5.**—Alignment of *H. hermesi* and GD1 orthologous pairs were used to determine (A) Ka, Ks and Ka/Ks ratio, (B) amino acid frequencies, and (C) codon frequencies. Significant difference between amino acid frequencies are indicated with an asterisk. All codon frequency differences were always significantly different between both species.
overrepresented pathways in GD1. Transcription factors were clearly more overrepresented in GD1: 71 transcription factors, based on BRITE protein families, and 20 basal transcription factors, based on KEGG pathways, were found in GD1 whereas 43 transcription factors (protein families) and 11 basal transcription factors were detected in *H. hermesi*. Moreover, of the shared genes most unigenes involved in transcription had higher gene expression values for GD1 compared with *H. hermesi*. This suggests that the activation and expression of transcription factors is much more prominent in species that inhabit fluctuating environments. Intertidal expression of transcription factors is much more prominent in *H. hermesi* compared with GD1. Significant differences (*P* < 0.05) are indicated with an asterisk. SDHB, succinate dehydrogenase (ubiquinone) iron-sulfur subunit; CYC1, ubiquinol-cytochrome c reductase cytochrome c1 subunit; QCR2, ubiquinol-cytochrome c reductase core subunit 2; ATPeV1C, V-type H+-transporting ATPase subunit C; COX1, cytochrome c oxidase subunit 1; NDUFA10, NADH dehydrogenase (ubiquinone) 1 alpha subcomplex subunit 10; ATPeF1A, F-type H+-transporting ATPase subunit alpha; ATPeV1H, V-type H+-transporting ATPase subunit H; NDUFV1, NADH dehydrogenase (ubiquinone) flavoprotein 1; NDUFB10, NADH dehydrogenase (ubiquinone) 1 beta subcomplex subunit 10.

**Oxidative Phosphorylation and Fatty Acid Metabolism Are More Strongly Expressed in Deep-Sea Environments**

Oxidative phosphorylation is the mitochondrial process in which the energy released by the transfer of electrons over electron transport carriers is used to establish a proton gradient which in turn is employed to produce energy under the form of ATP (Voet et al. 2006). Our results show that this process is more strongly expressed in *H. hermesi* and 43 *H. hermesi* unigenes were mapped against the oxidative phosphorylation compared with 35 mapped GD1 unigenes. Interestingly, we have identified a number of overexpressed genes, part of complex III (electron transport chain) in *H. hermesi*. The HMMV, the habitat of *H. hermesi*, is a chemosynthetic environment characterized by high hydrogen sulfide concentrations (Van Gaever et al. 2006). Hydrogen sulfide is a toxin that is capable of impairing biological processes in Metazoa. The toxicity of sulfide is mainly a result of inhibition of aerobic respiration by inhibition of cytochrome *c* oxidase (NRC 1979) and thus interfering with cellular respiration (Somero et al. 1989; Vismann 1991). Interestingly, the most overexpressed gene in *H. hermesi* involved in oxidative phosphorylation was cytochrome *c* oxidase *c* subunit I which might be a countermeasure to overcome sulfide poisoning. Sulfide can also be oxidized in the mitochondria and electrons can be incorporated into the electron transport chain by *ubiquinone* or *ubiquinol-cytochrome-c oxidoreductase* (complex III) (Powell and Somero 1986; Russell et al. 1989; Volkel and Grieshaber 1996). A higher expression of specific complex III genes might, therefore, be an important adaptation to sulfidic environments. However, oxidative phosphorylation is also influenced by a variety of other abiotic factors such as temperature (Toftaletti et al. 2003), pressure (Theron et al. 2000), salinity (Li et al. 2014), and oxygen levels (Mustroph et al. 2010). Oxygen levels rapidly deplete beneath 1 mm sediment depth at the HMMV (de Beer et al. 2006) while *H. hermesi* was...
identified up to 10 cm sediment depth. The effect of low oxygen levels include the induction of fermentation and glycolysis (Mustrph et al. 2010), influence mitochondria fusion (Tondera et al. 2009), mitochondrial biogenesis (Qin et al. 2014). and mitochondrial organelle number and volume (Gupta 2002), which implies that mitochondria are key players in the adaptation to low oxygen concentrations. Furthermore, the AMP-activated protein kinase signaling pathway, which is rapidly induced by low oxygen levels (Laderoute et al. 2006) was underrepresented in GD1 suggesting that oxygen levels are important. In summary, our results indicate that sulfide- and oxygen concentrations are presumably the most important environmental features explaining overrepresentation of genes involved in oxidative phosphorylation in Halomonhystera unigenes.

The deep-sea species further showed a clear response in fatty acid elongation compared with GD1. We observed an overrepresentation of the fatty acid elongation pathway whereas the fatty acid metabolism was underrepresented in GD1 both in KEGG and GO annotations. A recent study revealed that H. hermesi has higher proportions of highly unsaturated fatty acids (HUFAs, chain length ≥ 20 carbons and ≥ 3 double bond) due to increased proportions of eicosapentaenoic acid (EPA–20:5ω3) and the presence of docosahexaenoic acid (DHA–22:6ω3) which was absent in GD1 (Van Campenhout J and Vanreusel A submitted for publication). These HUFAs are believed to be essential in maintaining membrane fluidity under pressure (Yano et al. 1997) and docosahexaenoic acid (DHA–22:6ω3) was hypothesized to be the key to homeoviscous adaptations to depth and temperature changes in vertically migrating planktonic copepods (Pond et al. 2014). Given the absence of these HUFAs in the food source of H. hermesi (Van Gaever et al. 2009), the nematode has most likely the ability to synthesize these HUFAs. EPA, and DHA can be assimilated into triglyceride forms and stored until they are required for incorporation in the membrane, beta-oxidation, signaling activity, etc. (Williams and Burdge 2006; Riesner 2012). Our results at the molecular level now confirm that these HUFAs are indeed important to deep-sea adaptation.

GC Content, Amino Acid Composition, and Codon Usage Differs between Deep-Sea and Intertidal Species

Our results revealed that the genetic divergence (uncorrected p-distances) between orthologous pairs ranged from 8.3% to 40.1%. This observation further confirms the identity of both species and is influenced by the significant higher GC content in H. hermesi than in GD1 orthologous pairs. To date, many hypothesis such as directional mutation pressure (Sueoka 1988), metabolism (Martin 1995), CDS length (Oliver and Marin 1996), nitrogen fixation, genome size (Musto et al. 2006), DNA polymerase III subunit alpha (Zhao et al. 2007), and environmental pressure (Foerstner et al. 2005) have been postulated to explain varying GC contents. The GC content of both Halomonhystera species is correlated with codon usage bias, especially at the third codon position and with amino acid content, a pattern that has been frequently observed (Bernardi 1985; Wilquet and Van de Casteele 1999; Basak et al. 2004; Foerstner et al. 2005). Furthermore, nucleotide bias can have a dramatic effect on the amino acid composition of encoded proteins (Lory 1997; Singer and Hickey 2000) and it is currently believed that amino acid composition is more correlated to GC-content of the genome than to environment (Moura et al. 2013). In addition, arginine and alanine have been postulated as the “most piezophilic amino acids,” playing a role in adaptation to the piezophilic lifestyle in organisms (Di Giulio 2005; Jun et al. 2011; Pradel et al. 2013). One possible reason is the cost of ATP synthesis imposed on H. hermesi by low oxygen levels, low temperatures, and high sulfide concentrations. Whatever the driving force behind the GC content differences might be, it is clear that the different environments indirectly result in different codon usage with relatively small differences in amino acid abundances.

Few Gene Were Under Selection whereas Many Genes Were Uniquely Expressed

Adaptation depends on the natural selection of individuals with higher fitness thereby increasing their chance to produce more viable offspring (Darwin 1859). We could detect three genes that are subjected to positive selection. The 20S proteasome subunit beta 3 is part of the large 20S proteasome (28 subunits) the barrel shaped cylindrical core of the 26S proteasome, which degrades ubiquitinated proteins in an ATP-dependent process (Voet et al. 2006). This subunit has no catalytic activity (Ung et al. 2009) and its exact role in deep-sea adaptation remains unknown. The Splicing factor U2AF 65 kDa subunit is part of the U2 auxiliary factor important in mRNA splicing by recognizing the polypyrimidine tract at the 3’-end of introns (Wahl et al. 2009). U2AF65 preferentially binds on uridine-rich sites with frequent interruptions by cytidine residues but variation in the tract sequences and length affects the efficiency of splice site recognition (Singh et al. 1995). Halomonhystera hermesi has a higher GC content which might constrain the presence of uridine rich tracts in mRNA impairing the U2AF65 binding efficiency. Positive selection on U2AF65 might, therefore, be an indirect result of environmental factors acting on GC content and can be essential to better recognize splice sites. In addition, mechanisms such as alternative splicing are known to be a response to external stimuli and play an important role in gene function diversification and regulation (Stamm 2002; Ramani et al. 2011) but knowledge on its true role in adaptation remains scant.

The “myosin heavy chain” is involved in tight junctions and its amino acid composition can adaptively alter its pressure sensitivity (Morita 2008, 2010). Tight junctions,
underrepresented in GD1, establish physiological barriers that regulate the movement of ions, small solutes, and fluid between tissue compartments (Gonzalez-Mariscal et al. 2003). Both osmotic pressure and hydrostatic pressure are known to affect tight junctions (Tokuda et al. 2009; Gunzel and Yu 2013), suggesting that pressure and salinity may be important drivers for the adaptive changes in both species. In contrast to the low number of genes detected to be under positive selection, we observed a high rate of synonymous substitutions for those genes that were identified in both species. This highlights that new mutations which give rise to amino acid changes are removed from the populations by purifying selection, and that these genes are under high functional constraint. However, the molecular basis of adaptation is not only restricted to CDS but also includes noncoding regulatory sequences, for example, promoters (Lopez-Maury et al. 2008) and enhancers (Kamachi et al. 2009). Functionally important nontranslated sequences have been hypothesized to also be subject to positive and purifying selection (Andolfatto 2005). We currently lack information on these regulatory sequences and further research is required.

The low amount of genes under selection can be the result of several possibilities. First, increase in synonymous substitution rates may obscure detection of genes under positive selection (Erixon and Oxelman 2008). Second, it has been suggested that also synonymous sites are subject of nonneutral evolution (Chamary et al. 2006), but our analysis did not account for this. Third, we lack any information on adaptive evolution of uniquely expressed genes which might be under selection. Fourth, under the hypothesis that H. hermesi has an intertidal ancestor, this ancestor might have been preadapted to certain deep-sea conditions. Intertidal sediments, as well as decaying macroalgae, show similar environmental characteristics as a cold-seep environment such as hypoxic/anoxic conditions and high sulfide concentration as well as low temperatures in the winter and temporary high salinity levels are not out of the ordinary. As such, several genes might have been preadapted allowing synonymous mutations to accumulate.

Finally, many genes were uniquely expressed in both species. However, caution is required for the correct interpretation of our results as we here compared two species (without genomic data available) from two different environments and sampled only a single point in time. As such, it remains to be investigated whether the uniquely expressed genes are the result of adaptive evolution of, for example, regulatory sequences, or are in fact the result of environmental signals acting at the level of gene expression regulation at the time of sampling. Our results, combined with the tolerance of GD1 for balthyal cold-seep conditions (Van Campenhout et al. 2014), do suggest that plasticity should be taken into account. It remains to be determined to what extent plasticity or preadaptation of intertidal nematodes played a role in adaptive evolution of cold-seep nematodes.

Conclusion

In our study, the transcriptomes of two phylogenetically closely related marine nematode species from contrasting environments (deep sea and intertidal) have been compared. We revealed that most pathways are shared between both habitats, but several pathways and genes were differentially expressed in different habitats. Oxidative phosphorylation and fatty acid metabolism/elongation appeared to be of higher importance in the deep sea. In addition, we observed a higher GC content in H. hermesi resulting in higher amino acid frequencies with GC-rich codons affecting codon usage between H. hermesi and GD1 orthologous. Meanwhile the dynamic feature of GD1’s habitat was reflected in higher amounts of transcription factors, more strongly expressed signaling receptor activity, and a higher amount of active pathways and genes. Many genes were uniquely expressed in one species and only few genes show a signal that natural selection has been involved in their evolution. This indicates that Halomonhystera species might be preadapted to specific environmental conditions such as low oxygen levels and high sulfide concentrations.

Supplementary Material

Supplementary material is available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors thank chief scientist A. Boetius and project LOOME coordinator D. De Beer for the sampling opportunity during the RV Maria S. Merian campaign MSM 16/2. The authors acknowledge the captain and the crew of RV “Maria S. Merian” for their support with work at sea. The authors thank IFREMER (France), Woods hole oceanographic institution (MA, USA, AUV “Sentry”) and the Renard Centre of Marine Geology (Ghent University, Belgium, ROV “genesis”). Financial support was provided by the FP7 EU projects Hotspot Ecosystem Research and Man’s Impact On European Seas (Grant number 226354), Managing Impacts of Deep-sea resource exploitation (Grant number 603418) and European Sea Observatory NETwork (Grant number: 036851), and by the DFG for the Maria S. Merian cruise MSM 16/2.

Literature Cited

Basak S, Banerjee T, Gupta SK, Ghosh TC. 2004. Investigation on the causes of codon and amino acid usages variation between


Associate editor: Purificación López-García