SUPPORTING INFORMATION APPENDIX

TITLE

Stepwise Metamorphosis of the Tubeworm *Hydroides elegans* is Mediated by a Bacterial Inducer and MAPK Signaling

AUTHORS

Nicholas J. Shikuma[*,a,1,2], Igor Antoshechkin[*,a], João M. Medeiros[b], Martin Pilhofer[b], Dianne K. Newman[a,c]

*Contributed equally

AUTHOR AFFILIATIONS

[a]Division of Biology and Biological Engineering
California Institute of Technology
Pasadena, California 91125 USA

[b]Institute of Molecular Biology and Biophysics
Department of Biology, ETH Zürich
Otto-Stern-Weg 5, 8093 Zürich, Switzerland

[c]Howard Hughes Medical Institute
California Institute of Technology
Pasadena, California 91125 USA

CORRESPONDING AUTHOR AND PRESENT ADDRESS

[1]Nicholas J. Shikuma
[2]Department of Biology
San Diego State University
San Diego, California 92182 USA
nshikuma@mail.sdsu.edu

SI RESULTS

Overview of transcriptome dynamics of major life stages of *Hydroides*

The transition between larval and adult forms is a dramatic post-embryonic developmental change common to many marine invertebrates. We therefore determined the number of shared and uniquely expressed transcripts between the major life-stages. Surrounding a core set of expressed transcripts (39,219), *Hydroides* uniquely activates 4.7%, 16.8% and 19.4% of its transcriptome in pre-competent larvae, competent larvae and adult life stages, respectively (Figure S4).

Hydroides mitochondrial genome

As expected, the mitochondrial genome had a much higher median k-mer coverage than the nuclear genome and was assembled from the same set of reads in a separate velvet run with expected coverage and coverage cutoff set to 500 and 400, respectively. The longest contig of 18,604 bp was annotated using the MITOS web server (1) and found to encode the full complement of ribosomal and transfer RNAs, and ORFs except for atp8 (Figure S5, Dataset S3, S4). A graphical representation of the genome was generated by OrganellarGenomeDRAW (2).

SI METHODS

Genome sequencing and assembly

For genome sequencing, DNA was extracted from a single male worm using a Qiagen DNeasy blood and tissue kit. For each of two libraries with average insert sizes of 300 bp and 600 bp, 1 µg of genomic DNA was fragmented using Qsonica Q800R sonicator to generate the target size distribution. The fragmented DNA underwent end repair and dA tailing followed by adaptor ligation. Adaptor ligated DNA was size selected on agarose gel and sequences required for multiplexing and clustering were added by 6 cycles of PCR. PCR products were purified with Agencourt AMPure XP beads (Beckman Coulter). Libraries were quantified using Qubit fluorometer and the dsDNA HS Assay Kit (Invitrogen) and the size distribution was assessed with 2100 BioAnalyzer and the High Sensitivity DNA Kit (Agilent).

Velvet assembler (3) was run for a range of kmer sizes from 65 to 109. Quality of the assemblies was assessed based on N50, maximum contig length and assembly size metrics. The best-performing assemblies using Velvet assembler (3) were tested for completeness using the CEGMA pipeline (4). The best assembly was produced with kmer size of 101. It has a total size of 1.026 GB, maximum contig length of 244,066 bp and N50 of 17,328 bp. CEGMA pipeline estimates the completeness of this assembly at 89.11%. De novo repeat family identification was performed with RepeatModeler (5) and repeats were subsequently masked with RepeatMasker (6).

Transcriptome sequencing and analysis

mRNA was isolated from 1 µg of total RNA using NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB #E7490). RNA-seq libraries from larval samples were prepared with NEBNext Ultra RNA Library Prep kit (NEB #E7530) and the adult library was prepared using NEBNext Ultra Directional RNA Library Prep kit (NEB #E7420) according to manufacturer's instructions. Libraries were quantified with Qubit (Invitrogen) and the insert size distribution was assessed with 2100 BioAnalyzer (Agilent). All libraries were sequenced on Illumina HiSeq2500. Larval libraries were

sequenced in single read mode with the read length of 50 nt to the depth of 25 million reads per sample; the adult library was sequenced in paired end mode with the read length of 100 nt to the depth of 60 million reads. De novo gene prediction was carried out with AUGUSTUS (7). AUGUSTUS was first trained on a set of KOGs identified by CEGMA to produce Hydroides-specific HMM parameters, which were then used for gene prediction. RNA-seq libraries, both larval and adult, were aligned to the Hydroides genome using tophat2 (8). Alignment files were passed to cufflinks (9) for RNA-seq-based gene prediction. Gene models produced by AUGUSTUS and cufflinks were merged using the cuffmerge utility of the cufflinks package to generate the consensus gene set. To annotate predicted genes, transcript sequences were extracted and open reading frames were predicted using the TransDecoder utility of the Trinity software (10). Pfam domains were identified using HMMSCAN (11) and associated GO terms were added using Pfam2GO mapping supplied by InterPro (12). Differentially expressed genes were identified with DESeq R package (13). Briefly, gene models were quantified with htseq-count (14). For each time point and treatment, three biological replicates were used. Dispersions of the count data were estimated using 'per-condition' method and the sharingMode argument set to 'maximum'. Genes were considered differentially expressed if the multiple-testing-adjusted pvalue (padj) of the test was below 0.05 (FDR 5%). GO term enrichment analysis of differentially expressed gene sets was performed using the GOstats R package (15). Homologous genes in model animals were identified with Diamond (54) with an e-value cutoff of 0.001.

RNA isolation

RNAs were acquired by spawning independent groups of 3-5 males and 3-5 females and culturing larvae using the methods described previously (16, 17). Animals were harvested at 5 stages: (1) 4 days post fertilization (pre-competent), (2) 6 days post fertilization [competent, corresponding to larval stages 5 and 6, respectively (18)], larvae exposed to MACs for (3) 5 minutes and (4) 30 minutes, and (5) adults. Three biological replicate RNA samples were acquired for each larval stage of development. One sample was acquired for adults. For larvae exposed to MACs, a gentle MAC preparation was produced and larvae were exposed to a 100x dilution as described previously (17).

At the indicated developmental stages, portions of each larval culture were exposed to RNA preservation medium (20 mM EDTA, 25 mM sodium citrate, 5.3 M ammonium sulfate, pH 5.2) and stored at 4˚C for less than 24 h before extraction. RNA was extracted with a Qiagen RNeasy kit.

Metamorphosis assays

Stock solutions (50 mM) of U0126, SB203580 and SP600125 dissolved in DMSO were diluted in Instant Ocean artificial seawater. Competent larvae were exposed to DMSO only (0.02% V/V) or 0.25, 1, 2.5, 10 or 25 µM concentrations of each MAPK inhibitor for 2 hours before the addition of MAC extract. The percent of larvae that underwent metamorphosis was scored 24 hours after the induction of metamorphosis. Metamorphosis was scored visually by observing the number of individuals that formed branchial radioles, and a primary and secondary tube. Four biological replicates were performed for each treatment on two separate occasions with larvae spawned from different adults.

Bacterial strains, MAC extraction and quantification of MAC arrays

Bacterial mutant construction, MAC extraction and metamorphosis assays were described previously (17). Primers used are listed in Table S3. *P. luteoviolacea* HI1 Str^R wild type (19) or ∆12590-615 (strain no. NJS235) were modified to harbor *macB-sfgfp*, which fluorescently labels MAC arrays (17). MAC arrays from each strain were quantified using a haemocytometer and Leica inverted microscope. Three biological replicates with ≥6 images per replicate were acquired for each strain.

Vertical distribution assay

The vertical distribution assay was performed following Conzelmann et al. (20). Extracted MACs from wild type, ∆*macB* or ∆12590-615 strains were added to competent larvae in 1.5 ml capacity polystyrene cuvettes in 0.5 ml artificial seawater. The vertical distribution of larvae was imaged using a Keyence microscope (VHX-600) at the indicated time points. Twelve to 55 larvae were counted per replicate. Error bars indicate standard deviations of 6 biological replicates.

Quantitative Reverse Transcriptase PCR (qRT-PCR)

Larvae were exposed to MAC extract for 30 minutes and subsequently harvested for total RNA. For inhibitor experiments, larvae were exposed to SB203580 (10 µM) or SP600125 (10 µM) for 2 hours before adding MACs. RNA was extracted with a Qiagen RNeasy kit, 1 µg of total RNA was reverse transcribed with a Protoscript II cDNA Synthesis Kit (New England Biolabs) using oligo-dT primers and qRT-PCR reactions were performed using 2 to 8 µL of cDNA with iTaq Universal SYBR Green Supermix (Bio-Rad) on a real-time 7500 PCR Machine (Applied Biosystems). Primers used are listed in Table S3. Relative fold changes were calculated by comparing RNA samples from competent larvae. Four biological replicates were performed for each treatment on at least two separate occasions with larvae obtained from different adults.

Electron cryotomography

Electron cryotomography (ECT) was performed essentially as previously described (17). *P. luteoviolacea* was cultured (120 rpm, 30°C, 12 h) in 50 ml Marine Broth (Laboratorios CONDA). MACs were isolated and plunge-frozen as described previously (17). Images were collected using a 300 kV FEI Titan Krios transmission electron microscope equipped with a Falcon 2 direct electron detector. Tilt series were recorded with a 2° increment from -20° to +60° and from -22° to -60° with a 3.65 Å pixel size (wildtype), or with a 3° increment from -30° to +60° and -33° to -60° with a 2.3 Å pixel size (Δ12590-615). The cumulative dose was 95 e-/Å^2 and the defocus was set to -6 µm. Tilt series were collected with SerialEM (21). IMOD (22) was used for reconstruction and visualization.

Western blot analysis

Western blot analysis was performed as previously described (17). For experiments using inhibitors, competent larvae were exposed to SB203580 (10 µM) or SP600125 (10 µM) for 2 hours before the addition of MAC extract. No inhibitor and no MAC extract or MAC extract only treatments were performed as controls. Thirty minutes after the addition of MACs, larvae were resuspended and lysed in lysis buffer (50 mM Tris, pH

7.5, 150 mM NaCl, 0.1% Triton X-100, cOmplete EDTA-free protease inhibitor cocktail [Roche]) and proteins were subjected to western blot analyses. Antibodies used in this study are as follows: Phospho-p38 MAPK (4511), Phospho-SAPK/JNK (4668), Phospho-MAPKAPK-2 (3007), Phospho-ATF-2 (5112), Phospho-p53 (9284), Anti-rabbit IgG, HRP-linked (7074) antibodies (Cell Signaling Technology), p38alpha/beta (sc-7149), JNK (sc-571) antibodies (Santa Cruz Biotechnology) or GAPDH (600-401-A33) antibody (Rockland Antibodies and Assays). The top p38 (TCONS_00100902), JNK (TCONS_00114600) and MAPKAPK-2 (TCONS_00060448 - TCONS_00068388) homologs of *Hydroides* are 85%, 100% and 87% homologous, respectively, to the di-phosphorylated peptides used to produce the antibodies used in this work. Sequences for *Hydroides* ATF-2 and p53 covering the phosphorylated peptide sequences used for generating antibodies were not recovered in this genome assembly. A total of at least three biological replicates were performed for each treatment using larvae obtained from separate adults spawned on two separate occasions.

SI REFERENCES

1. Bernt M*, et al.* (2013) MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol* 69(2):313-319.
2. Lohse M, Drechsel O, Kahlau S, & Bock R (2013) OrganellarGenomeDRAW--a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res* 41(Web Server issue):W575-581.
3. Zerbino DR & Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18(5):821-829.
4. Parra G, Bradnam K, & Korf I (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23(9):1061-1067.
5. Smit A, Hubley, R. (2008) RepeatModeler Open-1.0.
6. Smit A, Hubley, R, Green, P (2013) RepeatMasker Open-4.0.
7. Hoff KJ & Stanke M (2013) WebAUGUSTUS--a web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic Acids Res* 41(Web Server issue):W123-128.
8. Kim D*, et al.* (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14(4):R36.
9. Trapnell C*, et al.* (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28(5):511-515.
10. Grabherr MG*, et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29(7):644-652.
11. Finn RD, Clements J, & Eddy SR (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 39(Web Server issue):W29-37.
12. Hunter S*, et al.* (2009) InterPro: the integrative protein signature database. *Nucleic Acids Res* 37(Database issue):D211-215.
13. Anders S & Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11(10):R106.
14. Anders S, Pyl PT, & Huber W (2015) HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31(2):166-169.
15. Falcon S & Gentleman R (2007) Using GOstats to test gene lists for GO term association. *Bioinformatics* 23(2):257-258.
16. Nedved BT & Hadfield MG (2009) *Hydroides elegans* (Annelida: Polychaeta): a model for biofouling research. *Marine and Industrial Biofouling* 4:203-217.
17. Shikuma NJ*, et al.* (2014) Marine tubeworm metamorphosis induced by arrays of bacterial phage tail-like structures. *Science* 343(6170):529-533.
18. Seaver EC, Thamm K, & Hill SD (2005) Growth patterns during segmentation in the two polychaete annelids, *Capitella* sp. I and *Hydroides elegans*: comparisons at distinct life history stages. *Evol Dev* 7(4):312-326.
19. Huang Y, Callahan S, & Hadfield MG (2012) Recruitment in the sea: bacterial genes required for inducing larval settlement in a polychaete worm. *Sci Rep* 2:228.
20. Conzelmann M*, et al.* (2011) Neuropeptides regulate swimming depth of *Platynereis* larvae. *Proc Natl Acad Sci U S A* 108(46):E1174-1183.

21. Mastronarde DN (2005) Automated electron microscope tomography using robust prediction of specimen movements. *Journal of Structural Biology* 152(1):36-51.
22. Kremer JR, Mastronarde DN, & McIntosh JR (1996) Computer visualization of three-dimensional image data using IMOD. *J Struct Biol* 116(1):71-76.

SI LEGENDS

Table S1. Repetitive sequence analysis of the *Hydroides* genome.

Table S2. Top Human homologs for three major subgroups of MAPK proteins (ERK, JNK and p38) and associated signaling components.

Table S3. Primers used in this work.

Table S4. Pfam domain frequency correlation. Pfam domain frequency correlation between *Hydroides elegans*, *Capitella teleta*, *Crassostrea gigas*, *Strongylocentrotus purpuratus*, *Homo sapiens*, *Caenorhabditis elegans*, *Drosophila melanogaster* and *Nematostella vectensis*.

Figure S1. Gene conservation analysis. 61540 proteins with BLASTP hits below the e-value cutoff of 1e-10 against the nonredundant database (NCBI, 11/15/2014) were classified based on their homology to genes in other major metazoan groups. **(A)** Homologs of 8842 proteins were found in all metazoan lineages (protostomia, deuterostomia, cnidaria and porifera), 14122 homologs exist in all eumetazoan clades (not in porifera), 17823 are specific to bilateria (protostomia, deuterostomia), 14001 are found in lophotrochozoa and deuterostomia, but not in ecdysozoa, while 1331 are protostome specific (present in lophotrochozoa and ecdysozoa, but not in deuterostomia). 5421 have homologs in lophotrochozoa only. **(B)** Number of *Hydroides* genes shared with major metazoan groupings based on their homology.

Figure S2. MACs stimulate signal transduction, tissue remodeling and innate immunity genes. Expression profiles of select (27%) differentially regulated genes before and after the stimulation of metamorphosis. The heatmap represents normalized expression values after $\log_2$ transformation. **(A)** Collagen domain-encoding and Matrix Metalloprotease (MMP) genes, **(B)** kazal-type Serine Protease Inhibitor (SPI) genes, **(C)** Von Willebrand Factor (VWF) domain-encoding genes, **(D)** ankyrin repeat domain-

encoding genes, **(E)** Epidermal Growth Factor (EGF) domain-encoding genes and **(F)** Mitogen-Activated Protein Kinase (MAPK)-associated genes.

Figure S3. MACs activate p38 and JNK MAPK pathways and expression of Fos and VWF genes. (A) Comparative western blot analysis of *Hydroides* homogenates against di-phosphorylated p38 or JNK or full length JNK. Levels of JNK protein were similar across treatments when probed with an antibody against the full-length human JNK protein. **(B)** Comparative western blot analysis of *Hydroides* homogenates against phosphorylation targets of p38 (phosphorylated MAPKAPK-2 [Thr334] or ATF-2 [Thr71]) or JNK (phosphorylated ATF-2 [Thr71] or p53 [Ser15]). Antibodies against the human p38 and GAPDH were not effective at detecting p38 or GAPDH from *Hydroides*. Expression of selected genes upon stimulation of metamorphosis in the presence or absence of **(C)** 10μM SB203580 or **(D)** 10μM SP600125 for 2 hours before the addition of MACs. RNA was harvested 30 minutes after MAC addition. Error bars indicate standard deviations of at least 3 biological replicates.

Figure S4. The number of shared and uniquely expressed transcripts (FPKM ≥ 1) between 4 day larva, 6 day larva and adult stages of *Hydroides* development. Circles and circle overlaps are plotted in proportion to the number of transcripts.

Figure S5. *Hydroides* mitochondrial genome encodes the full complement of ribosomal and transfer RNAs and all ORFs except for atp8. The inner circle represents GC content of the genome.

Movie S1. Timelapse microscopy of *Hydroides* metamorphosis in response to a bacterial biofilm of *P. luteoviolacea*. Time is hours:minutes:seconds.

Movie S2. Phenotypic response of *Hydroides* larvae to MACs extracted from wild type, Δ*12590-615* and Δ*macB* strains. Extract was added 5 seconds after the start of the movie. Time is hours:minutes:seconds.

Dataset S1. Genes differentially regulated between competent larvae and larvae exposed for 30 minutes to MACs or larvae exposed to MACs for 5 minutes and 30 minutes.

Dataset S2. Pfam domain list for *Hydroides* proteome.

Dataset S3. *Hydroides* mitochondrial genome sequence
(Shikuma_Antoshechkin_etal_Dataset_S3.fa).

Dataset S4. Genome annotations generated by MITOS
(Shikuma_Antoshechkin_etal_Dataset_S4.gff).

Table S1. Repetitive Sequence Analysis.

```
==================================================
file name: Helegans_assembly.k101.300min.fa
sequences:        188434
total length: 1026059160 bp  (1000953933 bp excl N/X-runs)
GC level:         35.47 %
bases masked:  348218675 bp ( 33.94 %)
==================================================
                number of      length   percentage
                elements*    occupied  of sequence
--------------------------------------------------
SINEs:             20637      7265068 bp    0.71 %
      ALUs           477       210281 bp    0.02 %
      MIRs          1099       153879 bp    0.01 %

LINEs:             89380     37853401 bp    3.69 %
      LINE1         6213      1911261 bp    0.19 %
      LINE2        21576      9365069 bp    0.91 %
      L3/CR1       17910      6703995 bp    0.65 %

LTR elements:      32811     12645391 bp    1.23 %
      ERVL             0            0 bp    0.00 %
      ERVL-MaLRs       0            0 bp    0.00 %
      ERV_classI    1169       380829 bp    0.04 %
      ERV_classII    596       316861 bp    0.03 %

DNA elements:     101015     27122440 bp    2.64 %
     hAT-Charlie    3791       605690 bp    0.06 %
     TcMar-Tigger      0            0 bp    0.00 %

Unclassified:     979530    250244881 bp   24.39 %

Total interspersed repeats:335131181 bp   32.66 %


Small RNA:          3137      1472579 bp    0.14 %

Satellites:         4115       892698 bp    0.09 %
Simple repeats:   208254     10742244 bp    1.05 %
Low complexity:    31658      1698982 bp    0.17 %
==================================================

* most repeats fragmented by insertions or deletions
  have been counted as one element
```

Table S2. MAPK signaling proteins in *Hydroides elegans*.

Pathway	Type	Human	Transcript No.	e-value	FPKM Pre-competent	FPKM Competent	FPKM adult
ERK	MAPK	ERK1	TCONS_00020254	2.00E-136	18.50	25.37	17.59
ERK	MAPK	ERK2	TCONS_00020254	8.00E-137	18.50	25.37	17.59
ERK	MAPKK	MEK1	TCONS_00004384	7.00E-82	0.01	0.87	0.11
ERK	MAPKK	MEK2	TCONS_00004384	3.00E-77	0.01	0.87	0.11
ERK	MAPKKK	RAF1	TCONS_00001283	0.00E+00	0.31	0.40	1.73
ERK	MAPKKK	BRAF	TCONS_00001285	0.00E+00	0.72	0.38	0.62
ERK	MAPKKK	MOS	TCONS_00131202	4.00E-53	0.00	0.02	0.00
JNK	MAPK	JNK1	TCONS_00120842	0.00E+00	5.86	8.82	4.78
JNK	MAPK	JNK2	TCONS_00120842	0.00E+00	5.86	8.82	4.78
JNK	MAPK	JNK3	TCONS_00120842	0.00E+00	5.86	8.82	4.78
JNK	MAPKK	MKK4	TCONS_00068499	4.00E-138	8.35	9.61	7.74
JNK	MAPKK	MKK7	TCONS_00007353	4.00E-97	1.45	2.16	0.82
JNK	MAPKKK	MEKK1	TCONS_00079769	6.00E-127	0.10	0.41	0.25
JNK	MAPKKK	MLK3	TCONS_00021065	7.00E-115	3.15	4.27	2.22
JNK	TF	p53	TCONS_00074557	1.00E-91	1.95	3.73	1.87
JNK/p38	TF	ATF-2	NH				
p38	MAPK	p38	TCONS_00100903	0.00E+00	5.47	8.44	7.44
p38	MAPK	p38B	TCONS_00100903	0.00E+00	5.47	8.44	7.44
p38	MAPK	p38delta	TCONS_00100903	2.00E-149	5.47	8.44	7.44
p38	MAPK	p38gamma	TCONS_00100903	2.00E-158	5.47	8.44	7.44
p38	MAPKK	MKK3	TCONS_00124993	9.00E-131	2.54	6.79	4.02
p38	MAPKK	MKK6	TCONS_00124993	9.00E-140	2.54	6.79	4.02
p38	MAPKKK	ASK1	TCONS_00118284	3.00E-111	2.20	5.48	1.76
p38	MAPKKK	ASK2	TCONS_00118284	2.00E-106	2.20	5.48	1.76
p38	MAPKKK	TAK1	TCONS_00038382	6.00E-168	1.02	1.92	0.63
p38	TF	MAPKAPK-2	TCONS_00060448-TCONS_00068388	2.00E-161	24.63	23.59	28.94

TF = Transcription factor.
NH = No Significant Hit. Gene not identified in current assembly.

Table S3. Primers used in this work.

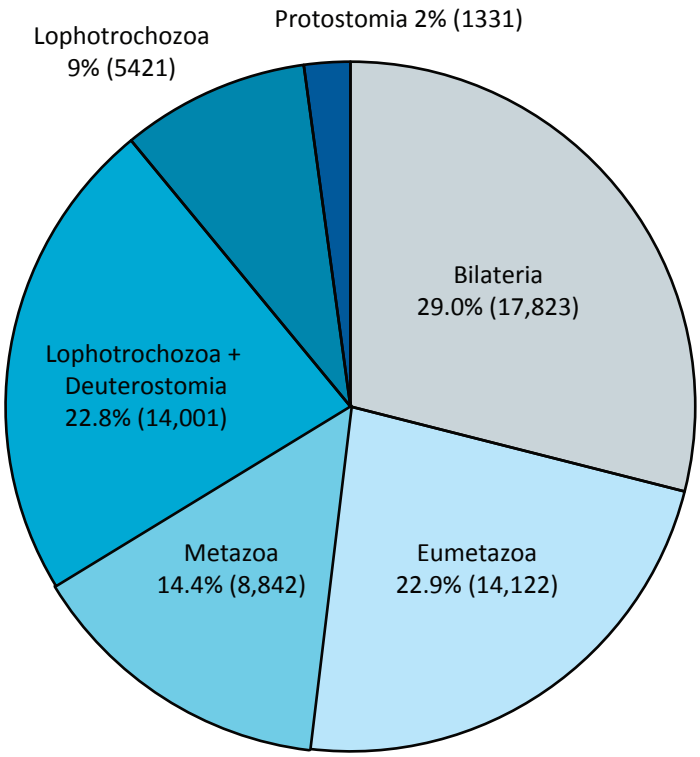
Primer	Sequence
Fos_qPCR_F	AATGCAGAAAGAGGCGAGAA
Fos_qPCR_R	CTGGCGCAGGTTGTAGATTT
VWF_qPCR_F	GTCTGGAAAACTTTGAGATGATAGTTGA
VWF_qPCR_R	AACCTCTTTCTCGAATTCCTCAAAATA
GAPDH_qPCR_F	TGTGGATCTGACCGTACGAC
GAPDH_qPCR_R	ACTCGAGAATGCCTTGCATC
12590-615_dA	GGTCGACGGATCCCAAGCTTCTTCTAGAGGTACCGCATGCTTGGAGCAATAAACGGGTTC
12590-615_dB	AACAGATCATTACATTAAAATGAGCCTAGCCATAAGGCCTCCTTGATA
12590-615_dC	TATCAAGGAGGCCTTATGGCTAGGCTCATTTTAATGTAATGATCTGTT
12590-615_dD	TTTTGAGACACAACGTGAATTCAAAGGGAGAGCTCCTTCTCCATTTTCGCCTTTG

Table S4. Pfam domain frequency correlation

	Hydroides	Capitella	C. gigas	S. purpuratus	Human	C. elegans	D. melanogaster	Nematostella
Hydroides	1.0000	0.8350	0.9038	0.8884	0.7855	0.6758	0.7873	0.8083
Capitella	0.8350	1.0000	0.8403	0.8874	0.7495	0.6806	0.7708	0.8942
C. gigas	0.9038	0.8403	1.0000	0.9006	0.8026	0.7224	0.8429	0.8299
S. purpuratus	0.8884	0.8874	0.9006	1.0000	0.7972	0.6932	0.8168	0.8744
Human	0.7855	0.7495	0.8026	0.7972	1.0000	0.6781	0.8851	0.7415
C. elegans	0.6758	0.6806	0.7224	0.6932	0.6781	1.0000	0.7697	0.7159
D. melanogaster	0.7873	0.7708	0.8429	0.8168	0.8851	0.7697	1.0000	0.7752
Nematostella	0.8083	0.8942	0.8299	0.8744	0.7415	0.7159	0.7752	1.0000

Figure S1

A


Lophotrochozoa
9% (5421)
Protostomia 2% (1331)
Bilateria
29.0% (17,823)
Lophotrochozoa +
Deuterostomia
22.8% (14,001)
Metazoa
14.4% (8,842)
Eumetazoa
22.9% (14,122)

B

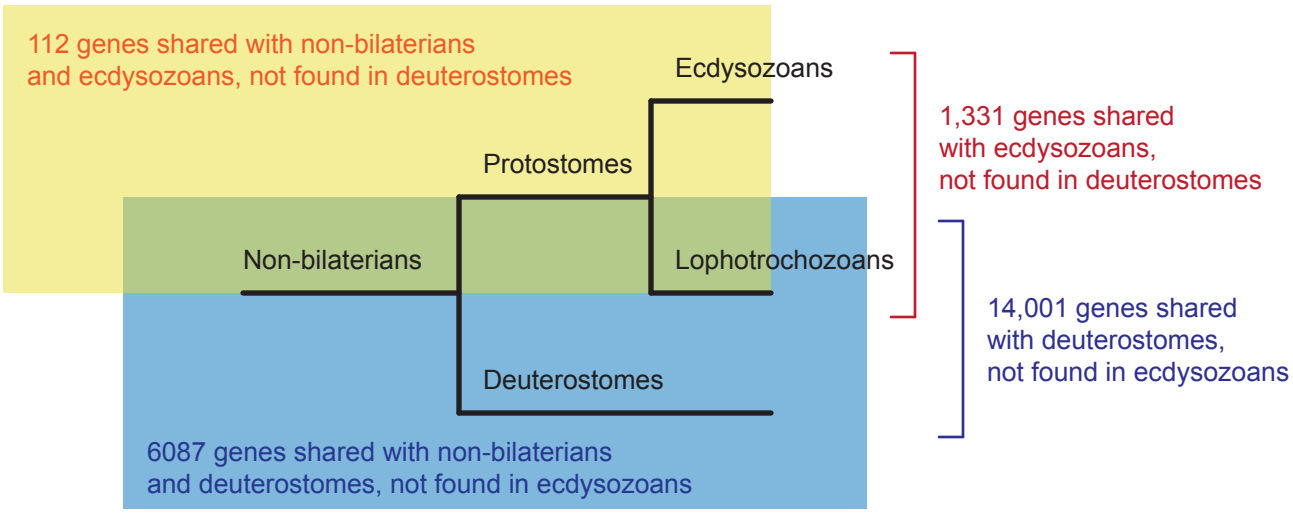

112 genes shared with non-bilaterians
and ecdysozoans, not found in deuterostomes
Ecdysozoans
Protostomes
Non-bilaterians
Lophotrochozoans
Deuterostomes
6087 genes shared with non-bilaterians
and deuterostomes, not found in ecdysozoans
1,331 genes shared
with ecdysozoans,
not found in deuterostomes
14,001 genes shared
with deuterostomes,
not found in ecdysozoans

Figure S2

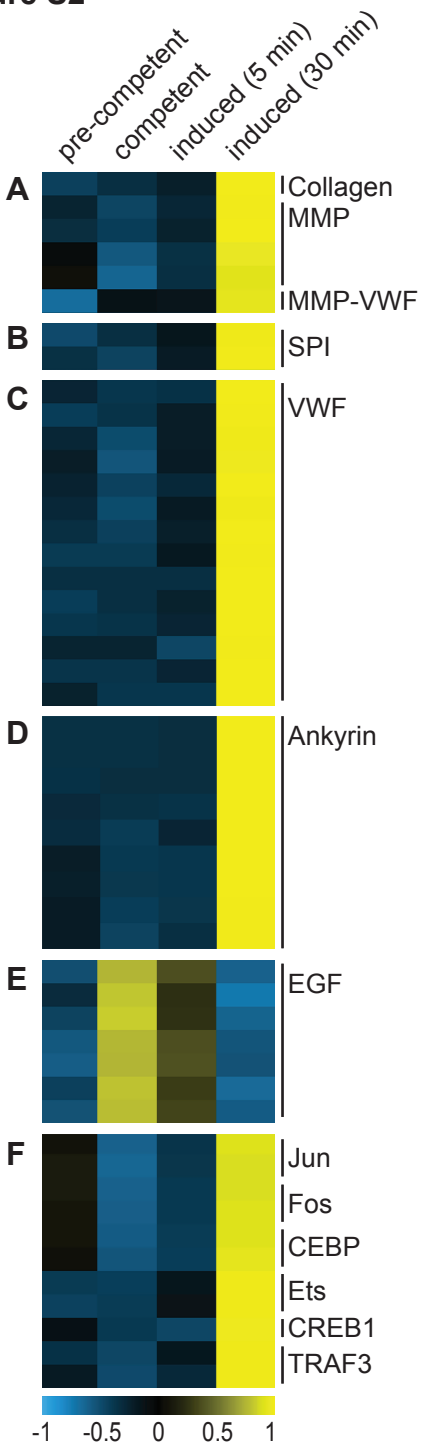

pre-competent
competent
induced (5 min)
induced (30 min)
A
Collagen
MMP
MMP-VWF
B
SPI
C
VWF
D
Ankyrin
E
EGF
F
Jun
Fos
CEBP
Ets
CREB1
TRAF3
-1
-0.5
0
0.5
1

Figure S3

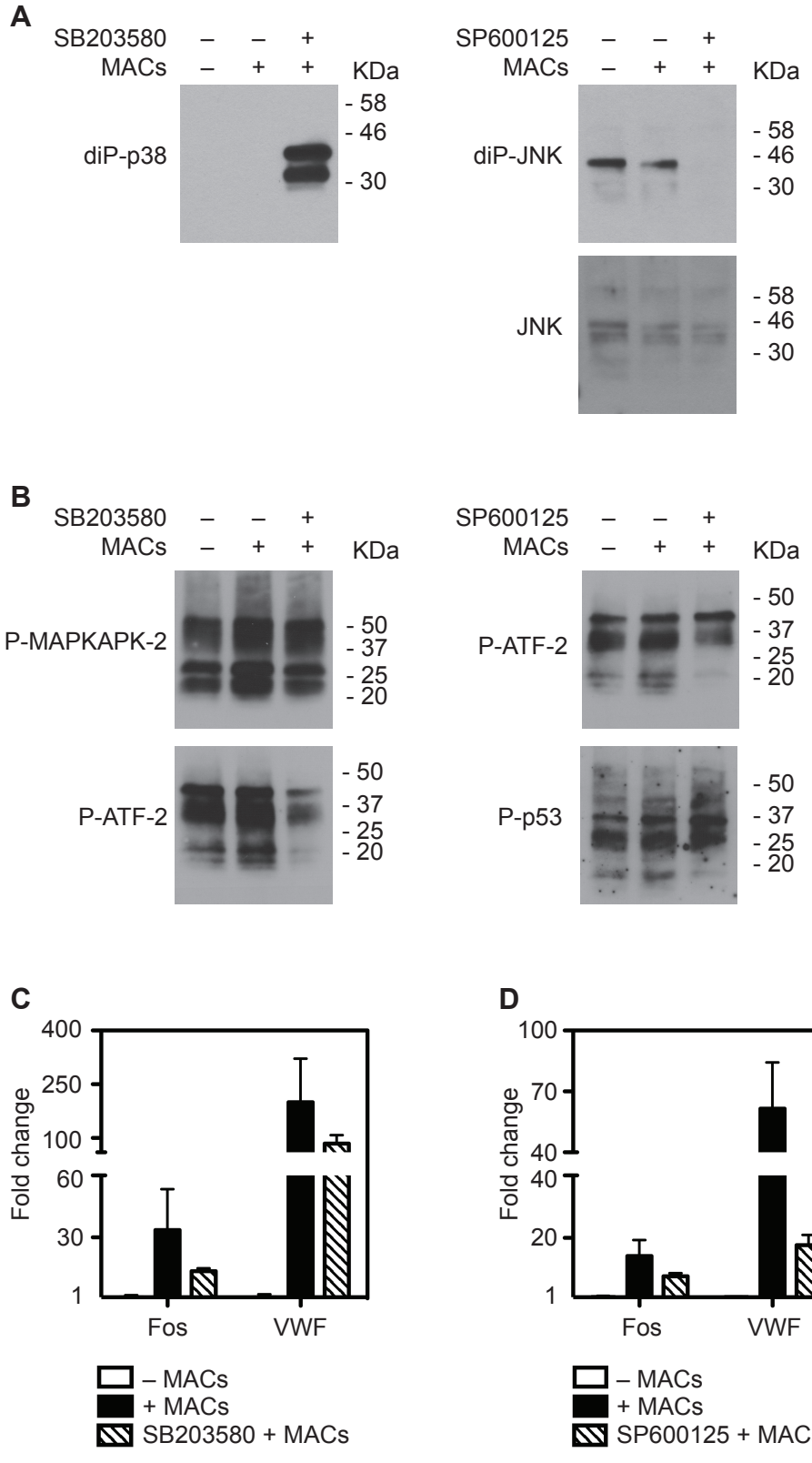

A
SB203580 – – +
MACs – + +
KDa
diP-p38
- 58
- 46
- 30
SP600125 – – +
MACs – + +
KDa
diP-JNK
JNK
- 58
- 46
- 30
B
SB203580 – – +
MACs – + +
KDa
P-MAPKAPK-2
P-ATF-2
- 50
- 37
- 25
- 20
SP600125 – – +
MACs – + +
KDa
P-ATF-2
P-p53
C
Fold change
400
250
100
60
30
1
Fos
VWF
– MACs
+ MACs
SB203580 + MACs
D
Fold change
100
70
40
40
20
1
Fos
VWF
– MACs
+ MACs
SP600125 + MACs

Figure S4

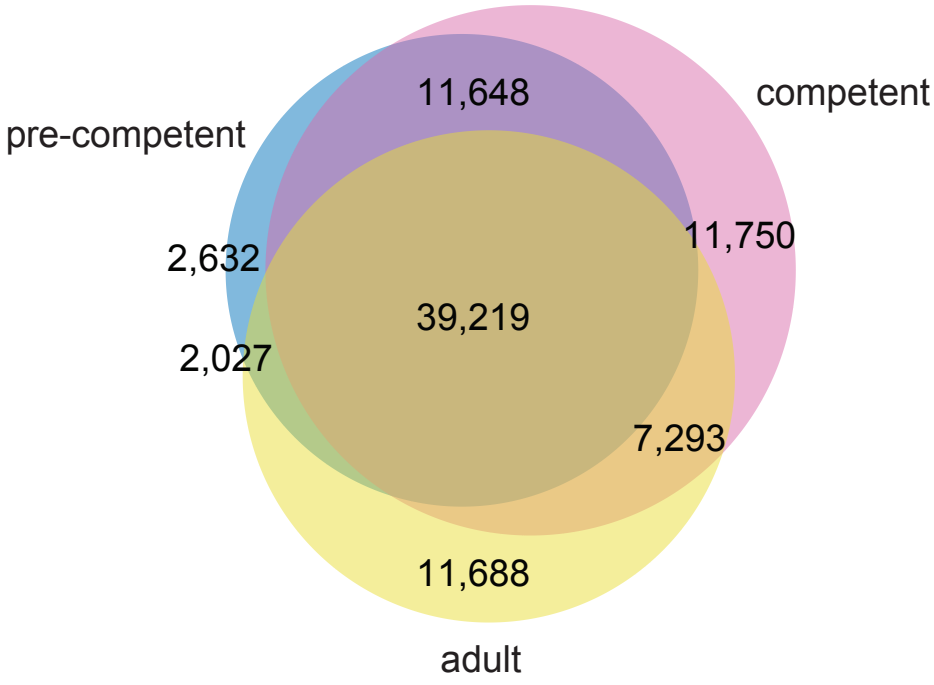
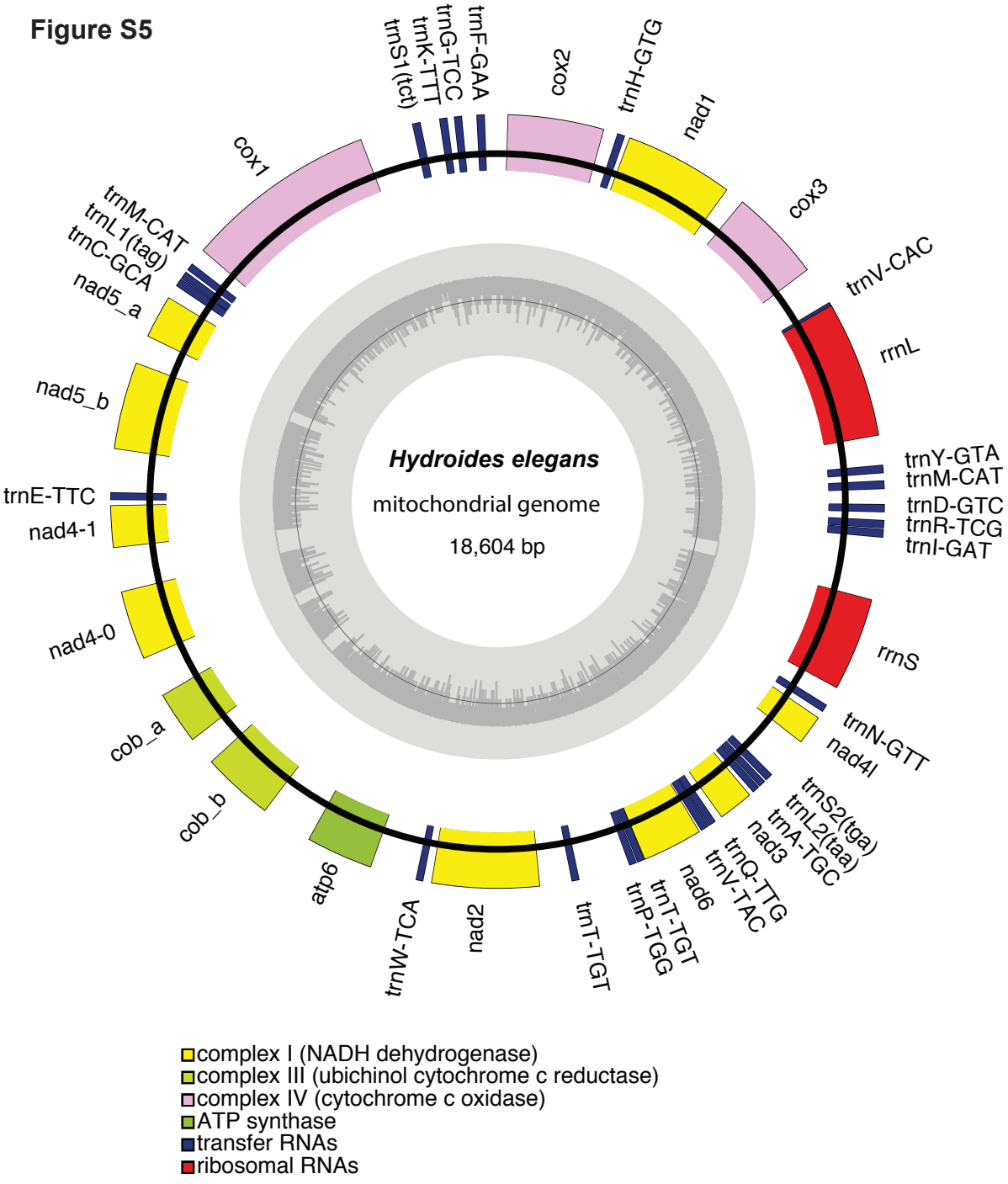

11,648
competent
pre-competent
11,750
2,632
39,219
2,027
7,293
11,688
adult

Figure S5


Hydroides elegans
mitochondrial genome
18,604 bp
trnF-GAA
trnG-TCC
trnK-TTT
trnS1(tct)
cox2
trnH-GTG
nad1
cox3
trnV-CAC
rrnL
trnY-GTA
trnM-CAT
trnD-GTC
trnR-TCG
trnI-GAT
rrnS
trnN-GTT
nad4l
trnS2(tga)
trnL2(taa)
trnA-TGC
nad3
trnQ-TTG
trnV-TAC
nad6
trnT-TGT
trnP-TGG
trnT-TGT
nad2
trnW-TCA
atp6
cob_b
cob_a
nad4-0
nad4-1
trnE-TTC
nad5_b
nad5_a
trnC-GCA
trnL1(tag)
trnM-CAT
cox1
complex I (NADH dehydrogenase)
complex III (ubichinol cytochrome c reductase)
complex IV (cytochrome c oxidase)
ATP synthase
transfer RNAs
ribosomal RNAs