Chromosome-wide evolution and sex determination in the three-sexed nematode *Auanema rhodensis*

Sophie Tandonnet\*, Georgios D. Koutsovoulos†,1, Sally Adams\*, Delphine Cloarec\*, Manish Parihar\*,2, Mark L. Blaxter†, Andre Pires-daSilva\*

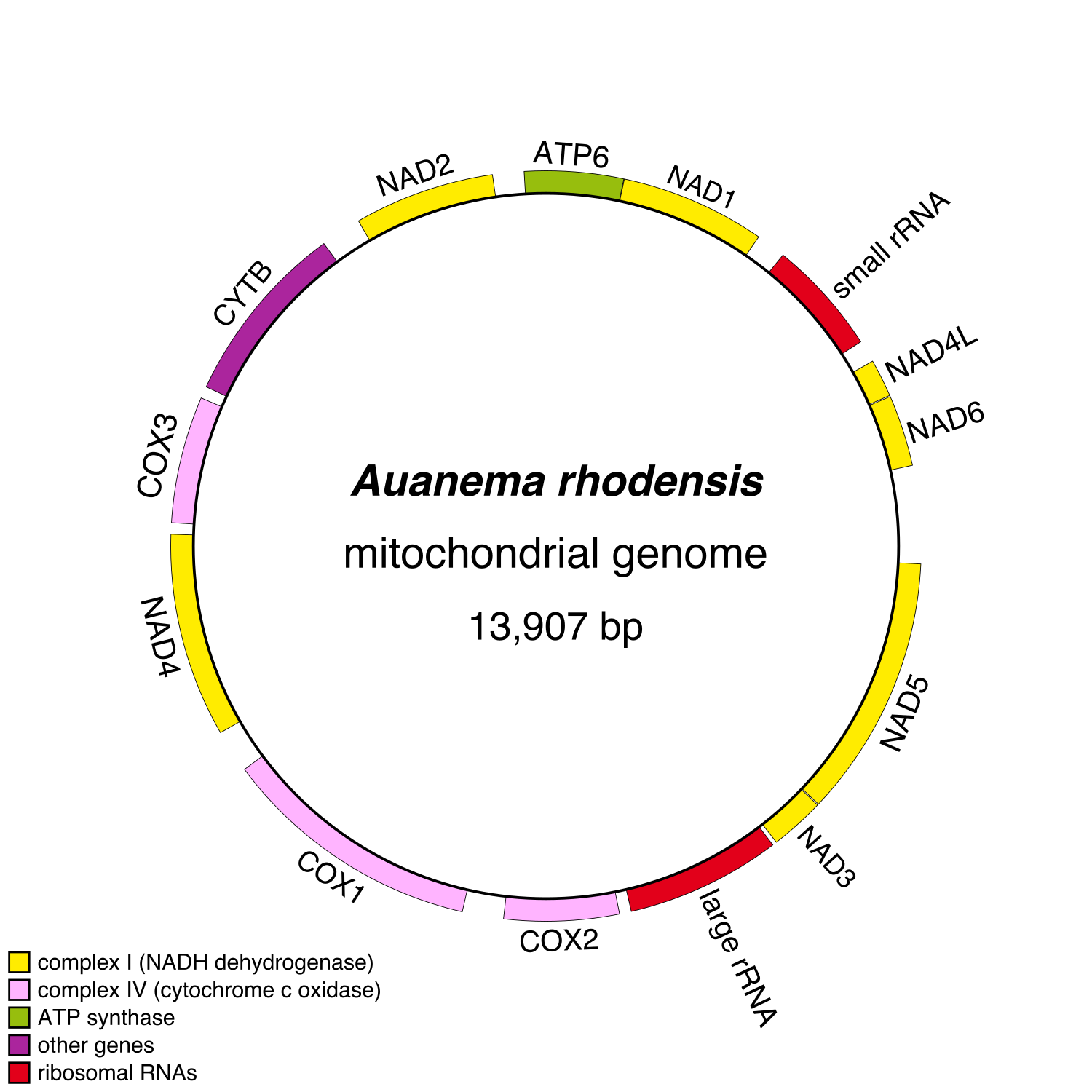
\**School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK*

†*Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK*

1*Present address: INRA, UMR1355 Institute Sophia Agrobiotech, F-06903 Sophia Antipolis, France*

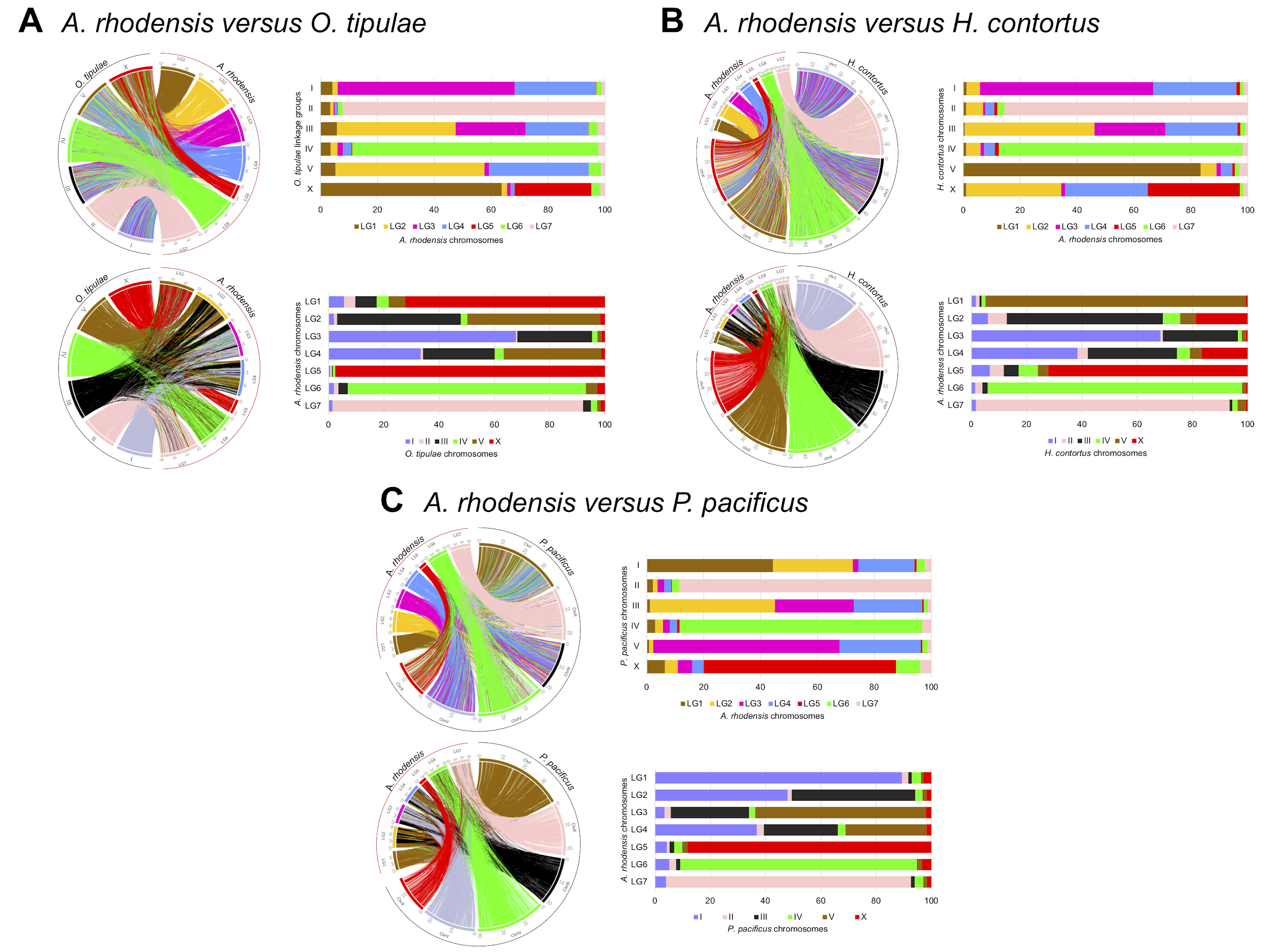
2*Present address: Department of Molecular Medicine, Institute of Biotechnology, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA.*

# Supplementary Figures



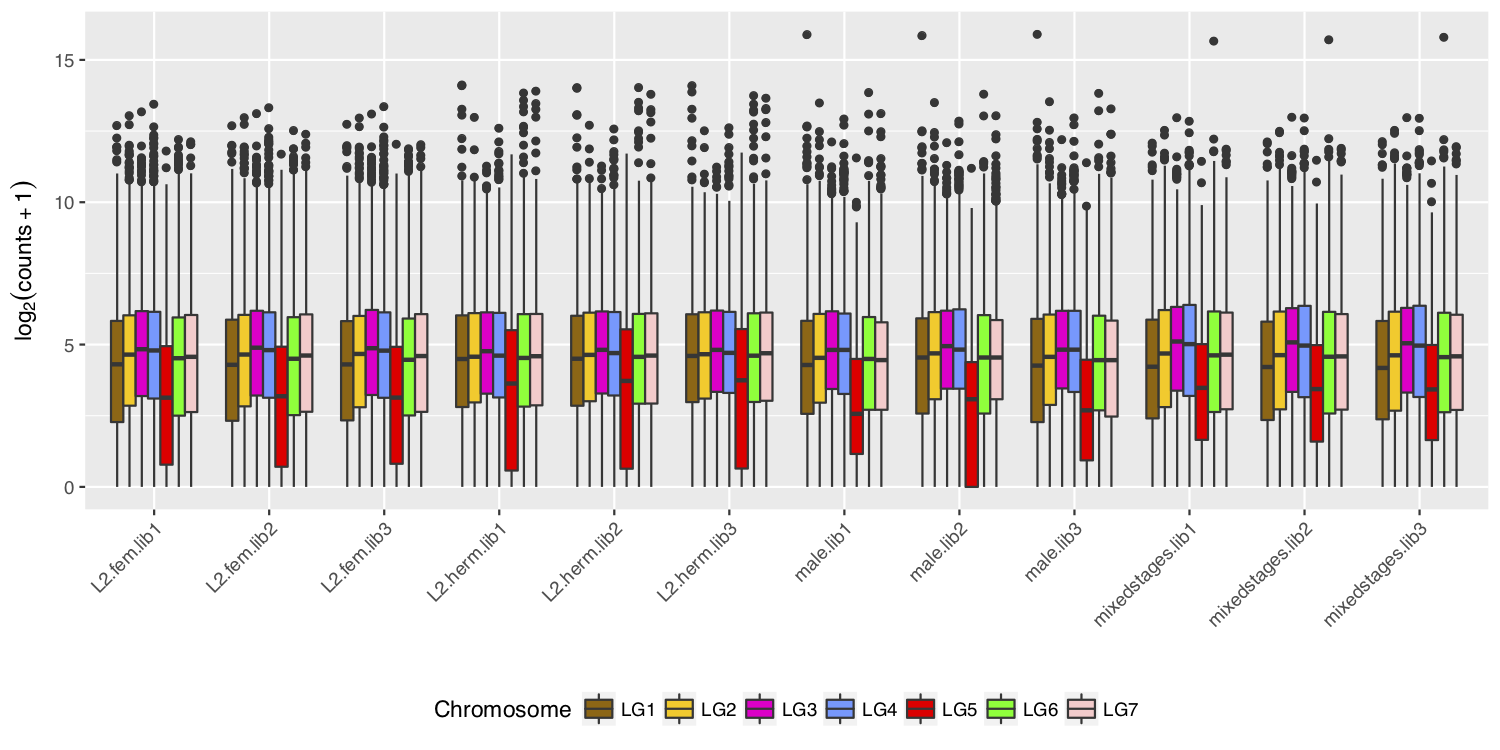
### Figure S1. The *Auanema rhodensis* mitochondrial genome.

The complete mitochondrial genome was found in “scaffold130\_61”. Manual inspection using the software Tablet (version 1.14.11.07, (Milne et al. 2013)), revealed an unjustified gap and an artefactual duplicated region in the scaffold. The mitochondrial genome was manually curated and annotated using Dogma (Wyman et al. 2004) and Geneious (version 6.1, (Kearse et al. 2012)). Both software packages annotated concordantly the 12 mitochondrial genes and the two ribosomal subunits. The two programs did not annotate concordantly the transfer RNAs (tRNAs) and therefore we did not include them in the current mitochondrial annotation. The corrected mitochondrial genome spans 13,907 bp, which is very similar to the *C. elegans* mitochondrial genome that is 13,794 nucleotides in length (Lemire 2005). The gene (and strand) collinearity is conserved between *C. elegans* and *A. rhodensis*.



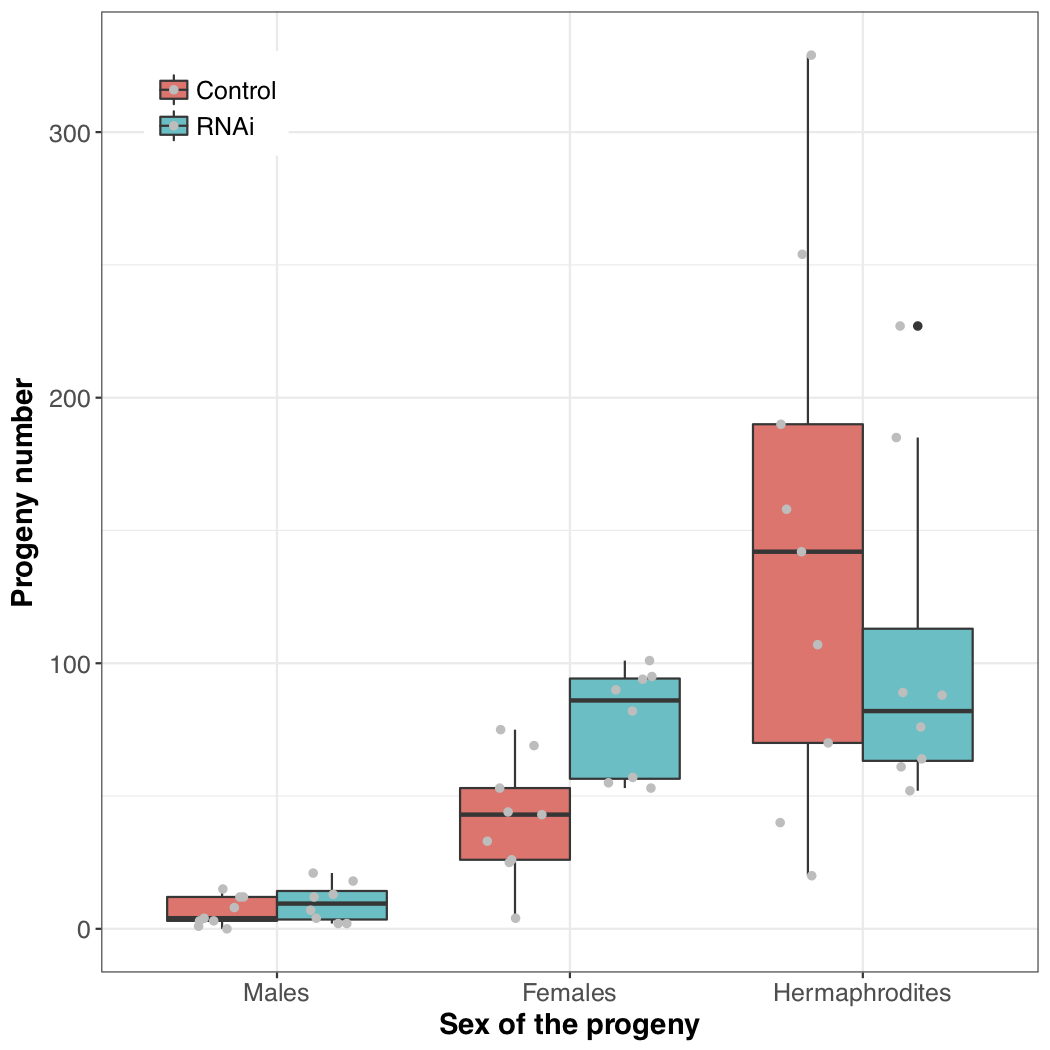
### Figure S2. Synteny relationships between chromosomes of *Auanema rhodensis* and other Rhabditid nematodes.

Location (Circos plots) and proportion (Bar plots) of the orthologous protein-coding genes between **(A)** *A. rhodensis and O. tipulae*, **(B)** *A. rhodensis and H. contortus* *and* **(C)** *A. rhodensis and P. pacificus*, coloured according to *A. rhodensis* chromosomes (top row of each panel) or to the compared species chromosomes (bottom row of each panel). Each line in the circos plots corresponds to a predicted orthologous gene between the two species.



### Figure S3. Expression of genes on the *A. rhodensis* X chromosome is generally lower than those on autosomes.

Boxplots of the log2 normalized expression of the genes located on each chromosome of *A. rhodensis* in different sexes, stages and libraries (biological replicates). The expression levels were normalized by the library size and log2-transformed for a better visualisation. The chromosome LG5 (in blue) is the X chromosome. "Fem", "Herm" and "lib" stand for female, hermaphrodite and library, respectively. This plot was generated using the R package ggplot2 (Wickham 2016).



### Figure S4. Boxplots of the number of progeny of each sex from control and RNAi injections targeting the DM domain transcription factor Arh.g5747 (*dmd-*10/11).

Each gray dot corresponds to an individual hermaphrodite mother.

## Supplementary Tables

### Table S1. Non-coding RNA loci in the genome of *Auanema rhodensis.*

|  |  |  |
| --- | --- | --- |
| **RNA gene type** | ***A. rhodensis*** | ***C. elegans estimates* (Stricklin et al. 2005)** |
| **Ribosomal RNAs** | | |
| **5S** | 45 | ~110 |
| **28S** | 10\* | ~55 |
| **18S** | 5\* | ~55 |
| **Transfer RNAs (tRNAs)** | 465 unique (495 non-unique) | 569 |
| **Spliced Leader RNAs** | 47 | 30 (10 SL1 and 20 SL2) |
| **Spliceosomal snRNAs (U1-6)** | 59  (19 U1, 9 U2, 4 U3, 14 U4, 7 U5, 6 U6) | 72  (12 U1, 19 U2, 5 U4, 13 U5, 23 U6) |
| **Signal Recognition Particle (SRP) RNA** | 12 | 5 |
| **Other small RNAs (microRNAs, snoRNAs, snRNAs)** | 145 | ~100 microRNAs  37 tncRNAs  25 snoRNAs |

siRNA, tncRNA, snoRNA, snRNA stand for small interfering RNA, tiny noncoding RNA, small nucleolar RNA, small nuclear RNA, respectively

\* Found by BLAST using the 18S and 28S partial sequences of *A. rhodensis* from GenBank

### 

### Table S2. Conserved genes between *A. rhodensis* and *D. melanogaster*.

Number of conserved genes between *A. rhodensis* and *D. melanogaster* and their proportion compared to non-conserved genes, by chromosome. The proportion of conserved genes on the *A. rhodensis* X chromosome is significantly smaller than on the autosomes (Z-Score -6.6676, p-value 0)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Linkage group*** | ***Assembly length*** | ***Number of genes*** | ***Number of conserved genes*** | ***Proportion of conserved genes*** |
| LG1 | 8489927 | 1,538 | 526 | 0.34 |
| LG2 | 9627060 | 1,760 | 745 | 0.42 |
| LG3 | 8741542 | 1,748 | 866 | 0.49 |
| LG4 | 8804062 | 1,586 | 735 | 0.46 |
| **LG5 (X)** | **3488253** | **604** | **169** | **0.27** |
| LG6 | 9421540 | 1,871 | 714 | 0.38 |
| LG7 | 9306279 | 1,754 | 692 | 0.39 |
| **All Autosomes** | **54390410** | **10257** | **4278** | **0.41** |

### Table S3. Results of the Gene Ontology (GO) analysis comparing the X chromosome GO categories to the autosomal ones.

‘U’ and ‘O’ refer to ‘under-representation’ and ‘over-representation’ of the GO term in the X compared to the autosomes. ‘BP’, ‘CC’ and ‘MF’ stand for ‘Biological Process’, ‘Cellular Component’ and ‘Molecular Function’, respectively. ‘Nr Annot’ and ‘Nr Non Annot’ refer to the number of genes annotated or non-annotated with that particular GO term.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| O/U | GO ID | GO Name | GO Category | FDR | P-Value | Nr Annot X | Nr Annot. Aut. | Nr Non Annot X | Nr Non Annot Aut. |
| U | GO:0006412 | translation | BP | 0.005 | 1.58E-05 | 0 | 204 | 604 | 10053 |
| O | GO:0007218 | neuropeptide signaling pathway | BP | 0.013 | 5.58E-05 | 8 | 18 | 596 | 10239 |
| U | GO:0005840 | ribosome | CC | 0.017 | 7.65E-05 | 0 | 173 | 604 | 10084 |
| U | GO:0003676 | nucleic acid binding | MF | 0.024 | 1.27E-04 | 33 | 1023 | 571 | 9234 |
| U | GO:0043231 | intracellular membrane- bounded organelle | CC | 0.025 | 1.35E-04 | 28 | 911 | 576 | 9346 |

### 

### 

### Table S4. Number of within- and between-strain variants per linkage group and their density (variants/bp).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Approximate Length** | **Intra-APS4** | **Intra-APS6** | **Inter-strains** | **Intra-APS4 variant density** | **Intra-APS6 variant density** | **Inter-strain variant density** |
| LG1 | 8489927 | 546 | 1185 | 26913 | 6.43E-05 | 1.40E-04 | 3.17E-03 |
| LG2 | 9627060 | 507 | 1281 | 17869 | 5.27E-05 | 1.33E-04 | 1.86E-03 |
| LG3 | 8741542 | 593 | 1571 | 19614 | 6.78E-05 | 1.80E-04 | 2.24E-03 |
| LG4 | 8804062 | 381 | 955 | 17894 | 4.33E-05 | 1.08E-04 | 2.03E-03 |
| **LG5 (X)** | **3488253** | **793** | **1475** | **11734** | **2.27E-04** | **4.23E-04** | **3.36E-03** |
| LG6 | 9421540 | 764 | 1653 | 19149 | 8.11E-05 | 1.75E-04 | 2.03E-03 |
| LG7 | 9306279 | 419 | 1022 | 19909 | 4.50E-05 | 1.10E-04 | 2.14E-03 |
| **Aut.** | **54390410** | **3210** | **7667** | **121348** | **5.90E-05** | **1.41E-04** | **2.23E-03** |

## 

### 

### 

### Table S5. Orthologs of known sex determination genes of *C. elegans*.

Through reciprocal BLASTp (see methods) we identified the potential orthologs of the main sex determination genes of *C. elegans* (Stothard and Pilgrim 2003). No clear orthologs were found for the sex determination genes *sex-1, xol-1, sdc-1, sdc-3, tra-2, fog-1, fog-2, fem-3, fbf-1, fbf-2, nos-3* in *A. rhodensis*. This could be due to the fast evolution of sex determination genes or to their absence.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ***A. rhodensis* locus** | **Chromosome** | **BLASTp E value (against *C. elegans*)** | **BLASTp E value (against *A. rhodensis*)** |
| ***fem-1*** | Arh.g7330 | LG6 | 9.66E-157 | 2.87E-144 |
| ***fem-2*** | Arh.g3906 | LG2 | 3.69E-74 | 1.60E-74 |
| ***fog-3*** | Arh.g2821 | LG4 | 8.18E-43 | 3.58E-45 |
| ***fox-1*** | Arh.g8527 | LG5 (X) | 1.42E-60 | 2.45E-55 |
| ***gld-1*** | Arh.g5696 | LG1 | 1.51E-112 | 1.29E-114 |
| ***her-1*** | Arh.g4307 | LG1 | 1.48E-19 | 6.06E-20 |
| ***laf-1*** | Arh.g2153 | LG2 | 0 | 0 |
| ***mab-3*** | Arh.g797 | LG7 | 3.14E-44 | 3.77E-45 |
| ***mag-1*** | Arh.g8665 | LG3 | 1.66e-94 | 6.95e-95 |
| ***mog-1*** | Arh.g2876 | LG4 | 0 | 0 |
| ***mog-4*** | Arh.g4360 | LG7 | 0 | 0 |
| ***mog-5*** | Arh.g9856 | LG1 | 0 | 0 |
| ***mog-6*** | Arh.g10475 | LG7 | 0 | 0 |
| ***sdc-2*** | Arh.g8906 | LG2 | 2.46E-135 | 1.12E-135 |
| ***tra-1*** | Arh.g4999 | LG2 | 4.28E-77 | 2.54E-76 |
| ***tra-3*** | Arh.g8312 | LG2 | 0 | 0 |

## 

**File S1.** Results of the Gene Ontology (GO) analysis of differentially expressed (DE) genes. The GO enrichment analyses were conducted on the "down-regulated" and "up-regulated" genes separately for each comparison, using two-tailed Fisher's exact test (FDR < 0.05). The abbreviations "fem", "herm" and "conv-fem" stand for female, hermaphrodite and converted females. The tags "OVER" and "UNDER" denote GO terms found over- or under-represented in the DE gene set (test set) compared to the non-DE genes (reference set). "Nr test" and "Nr reference" refer, respectively, to the number of DE genes and non-DE genes annotated with the particular GO term in the corresponding comparison. Likewise, "Non-Annot test" and "Non-Annot reference" refers to the number of genes not annotated with the particular GO term (in the corresponding comparison), in the test and reference sets, respectively.

**Literature cited:**

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung *et al.*, 2012 Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28 (12):1647-1649.

Lemire, B., 2005 Mitochondrial genetics. *WormBook*:1-10.

Milne, I., G. Stephen, M. Bayer, P.J. Cock, L. Pritchard *et al.*, 2013 Using Tablet for visual exploration of second-generation sequencing data. *Brief Bioinform* 14 (2):193-202.

Stothard, P., and D. Pilgrim, 2003 Sex-determination gene and pathway evolution in nematodes. *Bioessays* 25 (3):221-231.

Stricklin, S.L., S. Griffiths-Jones, and S.R. Eddy, 2005 *C. elegans* noncoding RNA genes. *WormBook*:1-7.

Wickham, H., 2016 *Ggplot2*. New York, NY: Springer Science+Business Media, LLC.

Wyman, S.K., R.K. Jansen, and J.L. Boore, 2004 Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20 (17):3252-3255.