Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/6/14/eaaz9808/DC1

Supplementary Materials for

Genome elimination mediated by gene expression from a selfish chromosome

Elena Dalla Benetta, Igor Antoshechkin, Ting Yang, Hoa Quang My Nguyen, Patrick M. Ferree*, Omar S. Akbari*

*Corresponding author. Email: oakbari@ucsd.edu (O.S.A.); pferree@kecksci.claremont.edu (P.M.F.)

Published 3 April 2020, *Sci. Adv.* **6**, eaaz9808 (2020)
DOI: 10.1126/sciadv.aaz9808

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/6/14/eaaz9808/DC1)

Supplementary figures:

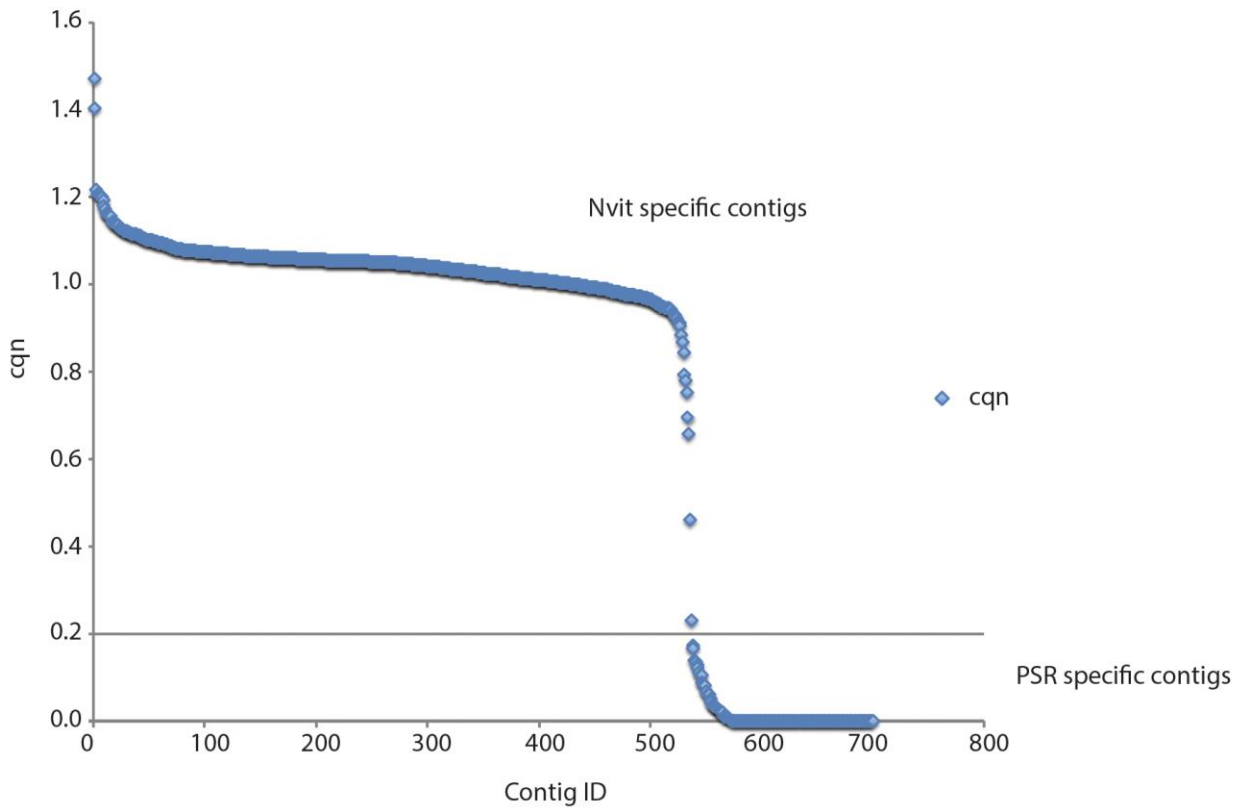


Fig. S1. Chromosome quotient of contigs

Chromosome quotient (cqn) represents a normalized ratio of the number of wild type and PSR+ reads that map to each contig. PSR-specific contigs are the ones with cqn smaller than 0.2.

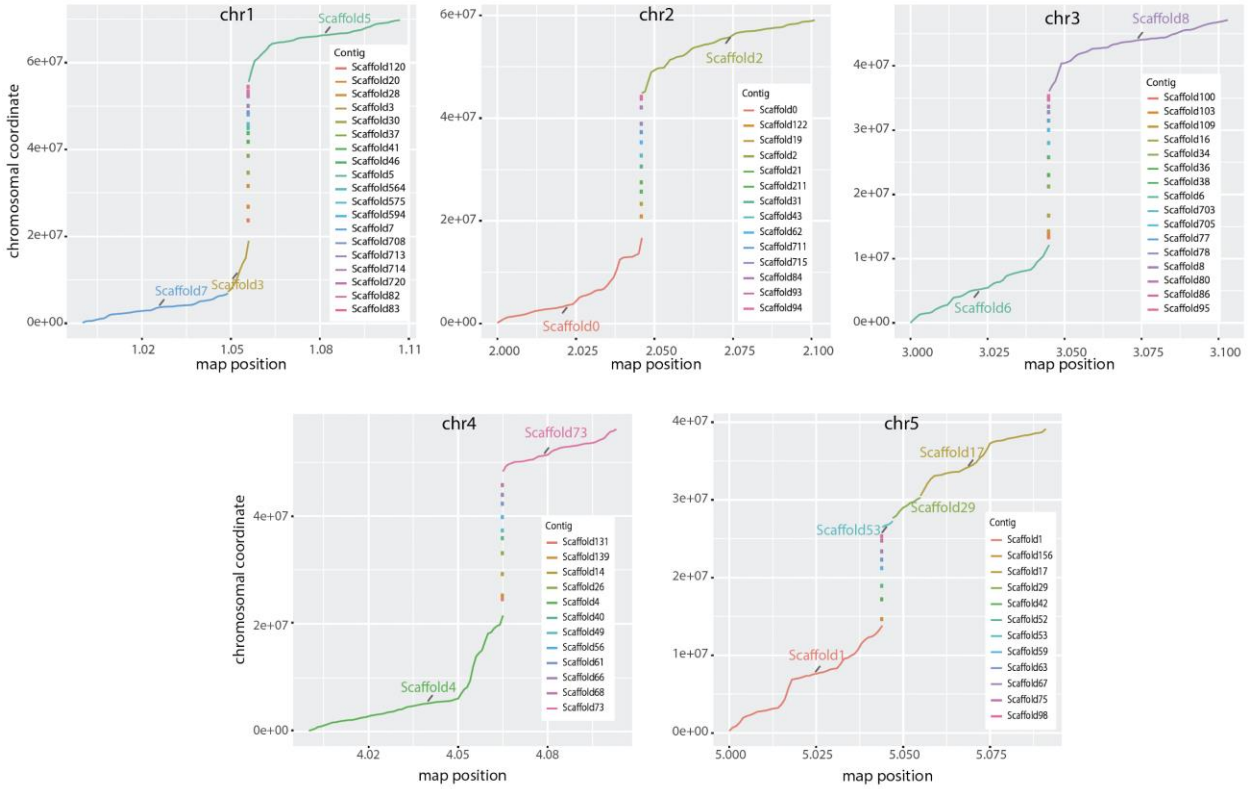


Fig. S2. Genetic map of *Nasonia vitripennis*

N. vitripennis contigs totaling 272 Mb (91.58% of total genome) were assigned to the five essential *N. vitripennis* chromosomes using the previously developed genetic markers (*22*).

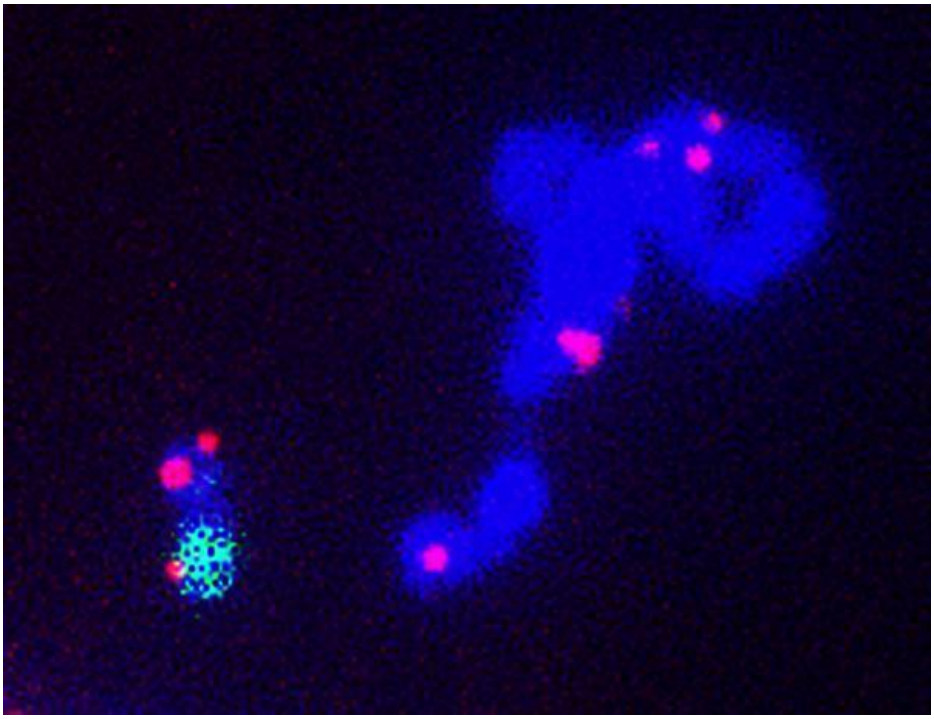


Fig. S3. Telomeres

Confocal imaging of Telomere sequences (TTATTGGG) on the 5 A-chromosome of *N. vitripennis* and on PSR chromosome. DNA is highlighted by DAPI (blue) and telomeres in red.

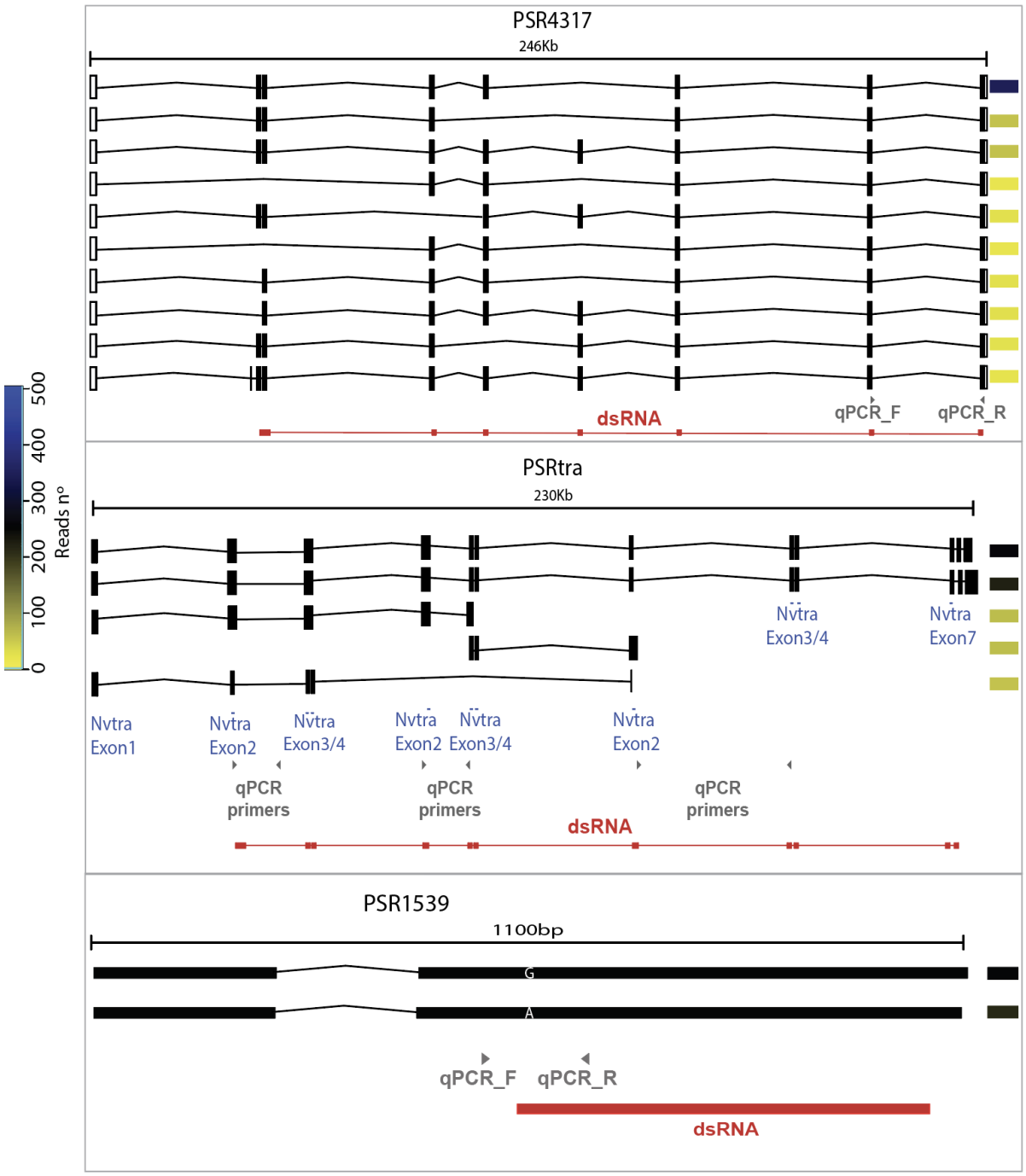
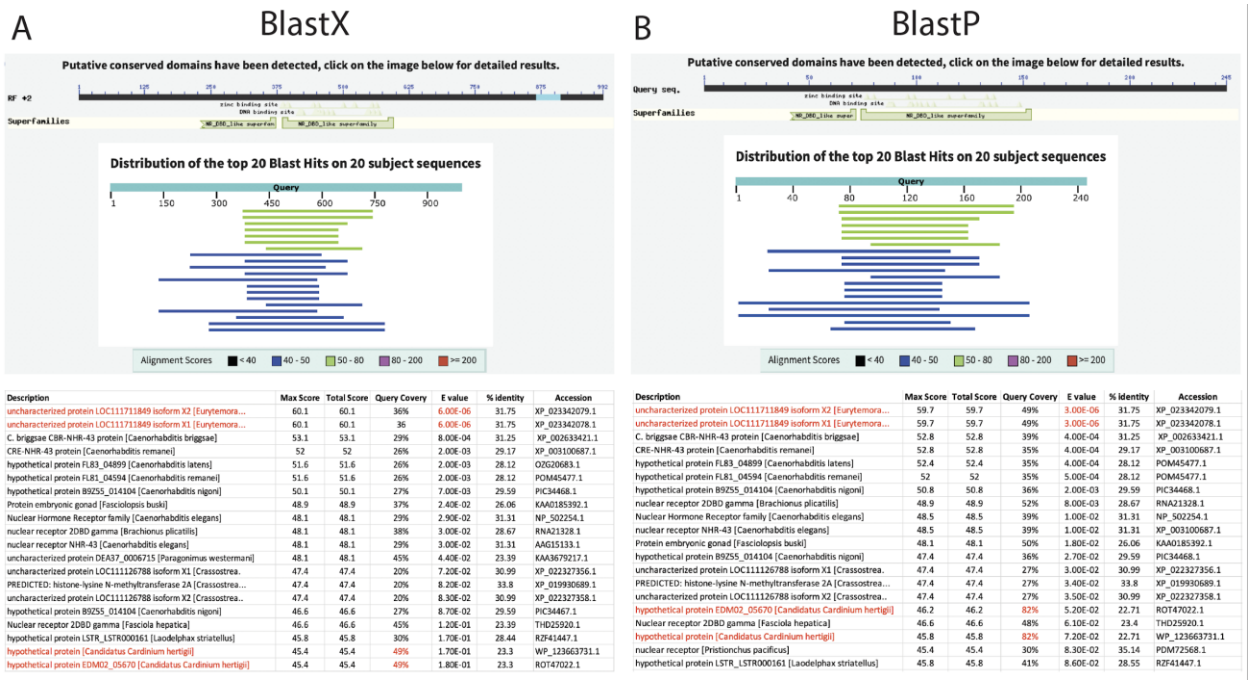


Fig. S4. gene model of candidate PSR genes

Schematic representation of five PSR candidate genes. Exons are indicated with boxes and intron with lines. Each row represents a splice variant and colored boxes on the right of each scheme indicate expression level in reads number (Yellow to blue for low to high expression value). The total length is indicated above each scheme in KB. Red lines indicate the region targeted by RNAi. Arrows indicate the qPCR primers location.



Description	Max Score	Total Score	Query Covery	E value	% identity	Accession
uncharacterized protein LOC111711849 isoform X2 [Eurytemora...	60.1	60.1	36%	6.00E-06	31.75	XP_023342079.1
uncharacterized protein LOC111711849 isoform X1 [Eurytemora...	60.1	60.1	36	6.00E-06	31.75	XP_023342078.1
C. briggsae CBR-NHR-43 protein [Caenorhabditis briggsae]	53.1	53.1	29%	8.00E-04	31.25	XP_002633421.1
CRE-NHR-43 protein [Caenorhabditis remanei]	52	52	26%	2.00E-03	29.17	XP_003100687.1
hypothetical protein FL83_04899 [Caenorhabditis latens]	51.6	51.6	26%	2.00E-03	28.12	OZG20683.1
hypothetical protein FL81_04594 [Caenorhabditis remanei]	51.6	51.6	26%	2.00E-03	28.12	POM45477.1
hypothetical protein B9Z55_014104 [Caenorhabditis nigoni]	50.1	50.1	27%	7.00E-03	29.59	PIC34468.1
Protein embryonic gonad [Fasciolopsis buski]	48.9	48.9	37%	2.40E-02	26.06	KAA0185392.1
Nuclear Hormone Receptor family [Caenorhabditis elegans]	48.1	48.1	29%	2.90E-02	31.31	NP_502254.1
nuclear receptor 2DBD gamma [Brachionus plicatilis]	48.1	48.1	38%	3.00E-02	28.67	RNA21328.1
nuclear receptor NHR-43 [Caenorhabditis elegans]	48.1	48.1	29%	3.00E-02	31.31	AAG15133.1
uncharacterized protein DEA37_0006715 [Paragonimus westermani]	48.1	48.1	45%	4.40E-02	23.39	KAA3679217.1
uncharacterized protein LOC111126788 isoform X1 [Crassostrea.	47.4	47.4	20%	7.20E-02	30.99	XP_022327356.1
PREDICTED: histone-lysine N-methyltransferase 2A [Crassostrea...	47.4	47.4	20%	8.20E-02	33.8	XP_019930689.1
uncharacterized protein LOC111126788 isoform X2 [Crassostrea..	47.4	47.4	20%	8.30E-02	30.99	XP_022327358.1
hypothetical protein B9Z55_014104 [Caenorhabditis nigoni]	46.6	46.6	27%	8.70E-02	29.59	PIC34467.1
Nuclear receptor 2DBD gamma [Fasciola hepatica]	46.6	46.6	45%	1.20E-01	23.39	THD25920.1
hypothetical protein LSTR_LSTR000161 [Laodelphax striatellus]	45.8	45.8	30%	1.70E-01	28.44	RZF41447.1
hypothetical protein [Candidatus Cardinium hertigii]	45.4	45.4	49%	1.70E-01	23.3	WP_123663731.1
hypothetical protein EDM02_05670 [Candidatus Cardinium hertigii]	45.4	45.4	49%	1.80E-01	23.3	ROT47022.1

Description	Max Score	Total Score	Query Covery	E value	% identity	Accession
uncharacterized protein LOC111711849 isoform X2 [Eurytemora...	59.7	59.7	49%	3.00E-06	31.75	XP_023342079.1
uncharacterized protein LOC111711849 isoform X1 [Eurytemora...	59.7	59.7	49%	3.00E-06	31.75	XP_023342078.1
C. briggsae CBR-NHR-43 protein [Caenorhabditis briggsae]	52.8	52.8	39%	4.00E-04	31.25	XP_002633421.1
CRE-NHR-43 protein [Caenorhabditis remanei]	52.8	52.8	35%	4.00E-04	29.17	XP_003100687.1
hypothetical protein FL83_04899 [Caenorhabditis latens]	52.4	52.4	35%	4.00E-04	28.12	POM45477.1
hypothetical protein FL81_04594 [Caenorhabditis remanei]	52	52	35%	5.00E-04	28.12	POM45477.1
hypothetical protein B9Z55_014104 [Caenorhabditis nigoni]	50.8	50.8	36%	2.00E-03	29.59	PIC34468.1
nuclear receptor 2DBD gamma [Brachionus plicatilis]	48.9	48.9	52%	8.00E-03	28.67	RNA21328.1
Nuclear Hormone Receptor family [Caenorhabditis elegans]	48.5	48.5	39%	1.00E-02	31.31	NP_502254.1
nuclear receptor NHR-43 [Caenorhabditis elegans]	48.5	48.5	39%	1.00E-02	31.31	XP_003100687.1
Protein embryonic gonad [Fasciolopsis buski]	48.1	48.1	50%	1.80E-02	26.06	KAA0185392.1
hypothetical protein B9Z55_014104 [Caenorhabditis nigoni]	47.4	47.4	36%	2.70E-02	29.59	PIC34468.1
uncharacterized protein LOC111126788 isoform X1 [Crassostrea.	47.4	47.4	27%	3.00E-02	30.99	XP_022327356.1
PREDICTED: histone-lysine N-methyltransferase 2A [Crassostrea...	47.4	47.4	27%	3.40E-02	33.8	XP_019930689.1
uncharacterized protein LOC111126788 isoform X2 [Crassostrea..	47.4	47.4	27%	3.50E-02	30.99	XP_022327358.1
hypothetical protein EDM02_05670 [Candidatus Cardinium hertigii]	46.2	46.2	82%	5.20E-02	22.71	ROT47022.1
Nuclear receptor 2DBD gamma [Fasciola hepatica]	46.6	46.6	48%	6.10E-02	23.4	THD25920.1
hypothetical protein [Candidatus Cardinium hertigii]	45.8	45.8	82%	7.20E-02	22.71	WP_123663731.1
nuclear receptor [Pristionchus pacificus]	45.4	45.4	30%	8.30E-02	35.14	PDM72568.1
hypothetical protein LSTR_LSTR000161 [Laodelphax striatellus]	45.8	45.8	41%	8.60E-02	28.55	RZF41447.1

Fig. S5. *haploidizer* Blast results

A) Graphic alignment and list of the top 20 blastx hit of *haploidizer* transcript. B) Graphic alignment and list of the top 20 blastp hit of *haploidizer* predicted ORF (Open Reading Frame). The order of the sequences is based on E value. Red values indicate the best hit based on E-value and query coverage.

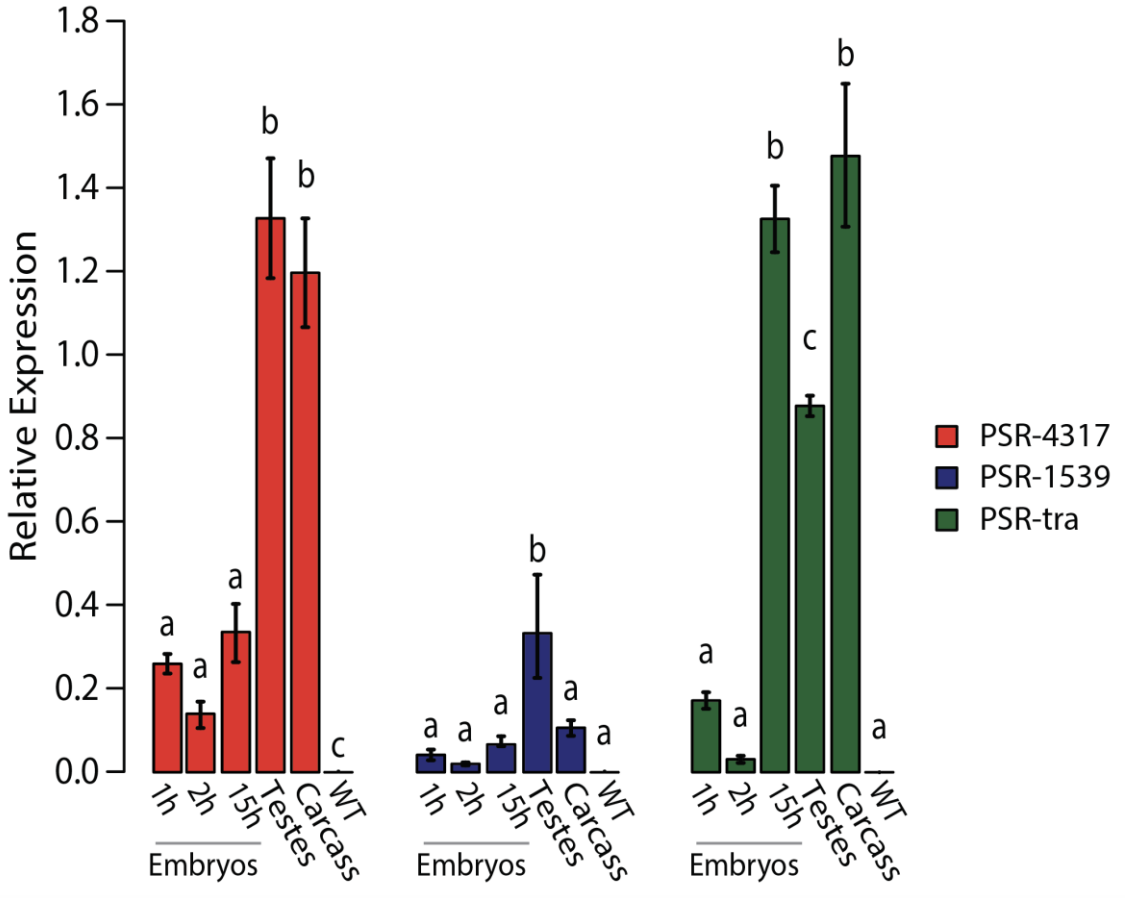


Fig. S6. Validation of expression of PSR candidate genes

Relative expression from RT-qPCR of PSR candidate genes among three embryonic time points, testes and carcass samples. Different letters indicate significance differences per tested gene between time points (n=5, $P < 0.05$, one-way ANOVA with Tukey's multiple-comparisons test).

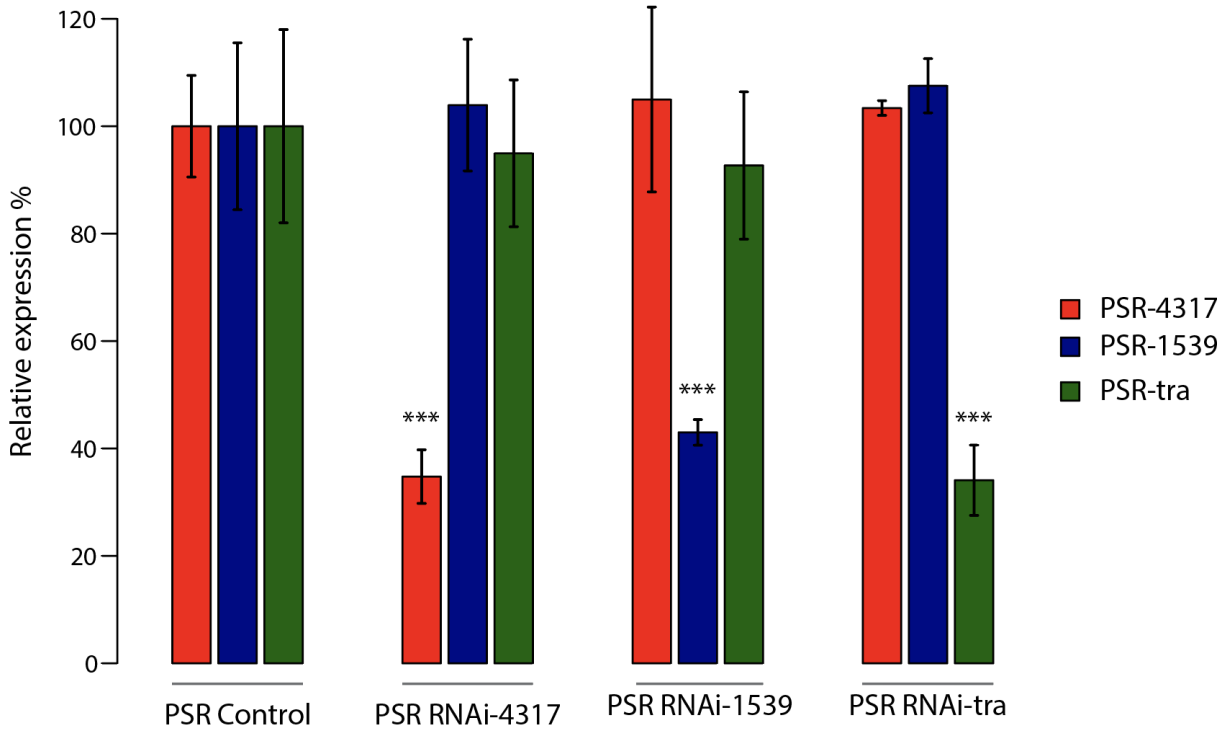


Fig. S7. RNAi efficiency

Relative gene expression of PSR candidate genes in untreated (PSR Control) and RNAi treated PSR males (PSR RNAi-4317, PSR RNAi-1539, PSR RNAi-tra) 24h after emergence. Different colours indicate different gene tested per each RNAi experiment. Asterisks indicate significant difference (n=5, $P < 0.05$, one-way ANOVA with Tukey's multiple-comparisons test).

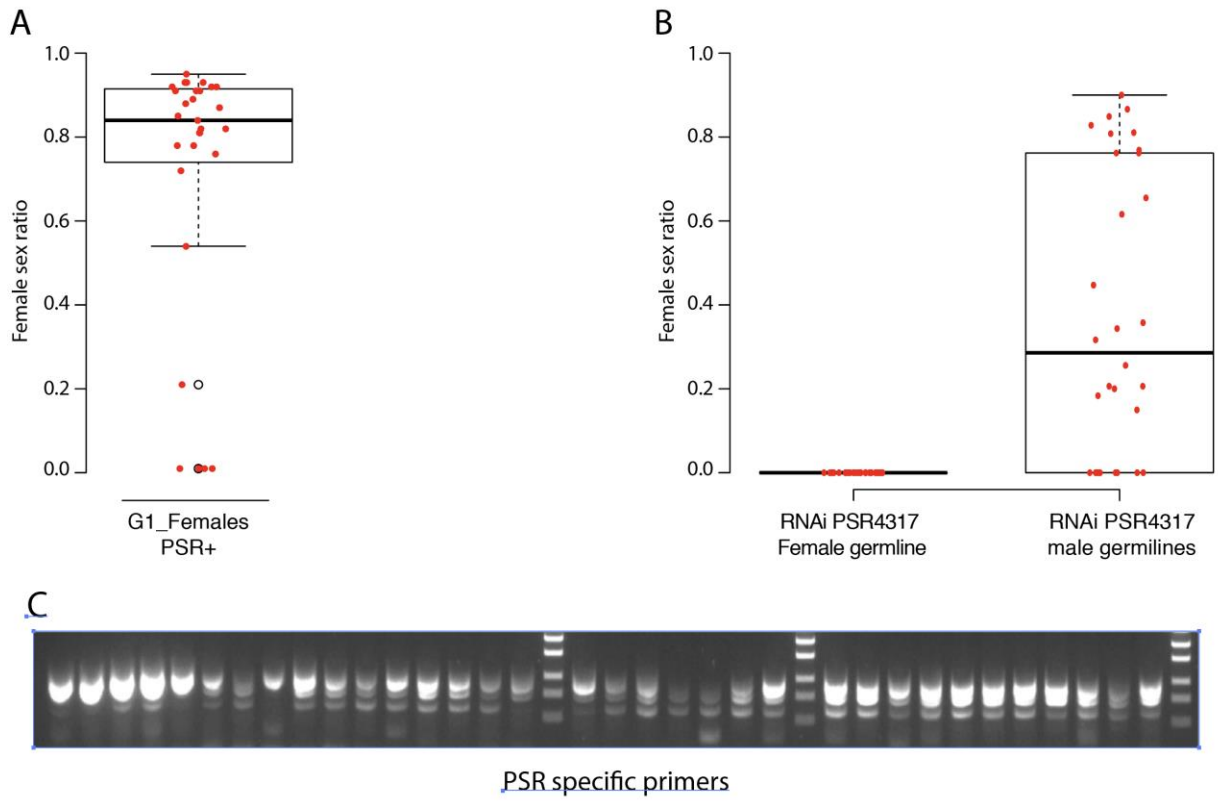


Fig. S8. Proportion of female offspring after PSR-RNAi

A) Box plot displaying the proportion of females produced from PSR+ G1 females originate from positive RNAi males. **B)** Box plot displaying the proportion of females after RNAi either in male or female germlines. **C)** Gel electrophoresis showing that tested G1 females carry PSR using PSR specific primers. PSR gene analysed using PSR specific primers (Table S10).

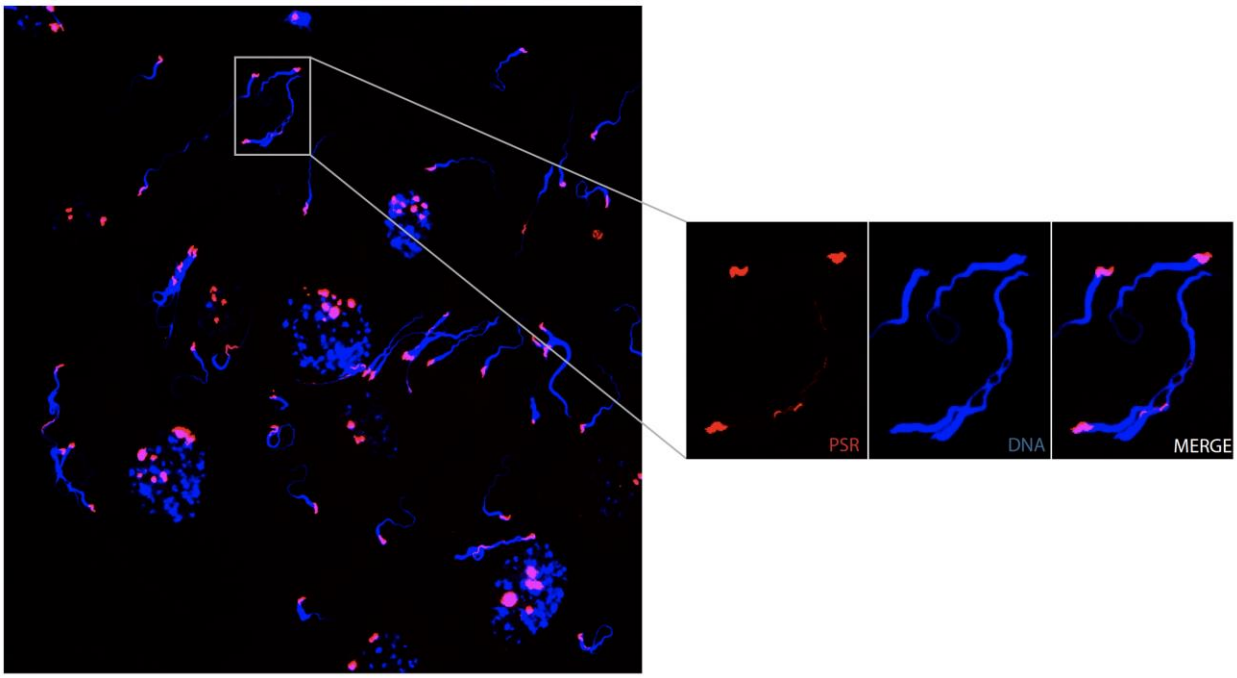


Fig. S9: Sperm fluorescent *in situ* hybridization

Fluorescent *in situ* hybridization of PSR+ RNAi-treated sperms. DNA is highlighted by DAPI (blue) and PSR by a sequence-specific FISH probe (red).

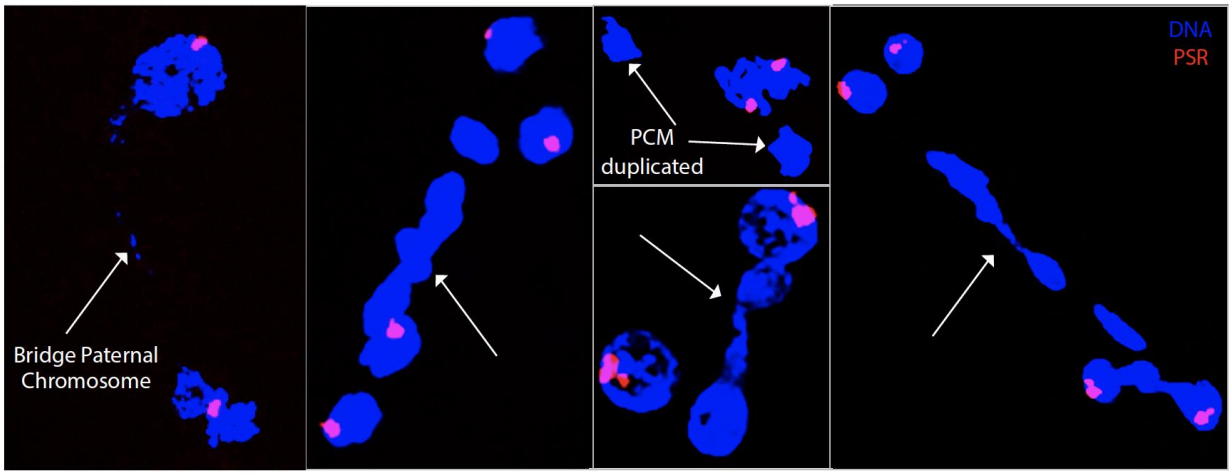


Fig. S10. Embryonic fluorescent *in situ* hybridization of RNAi-treated PSR+

Additional examples of Fluorescent *in situ* hybridization of RNAi treated PSR+ males. DNA is highlighted by DAPI (blue) and PSR by a sequence-specific FISH probe (red). White arrows indicated disrupted PCM (Paternal chromatin mass).

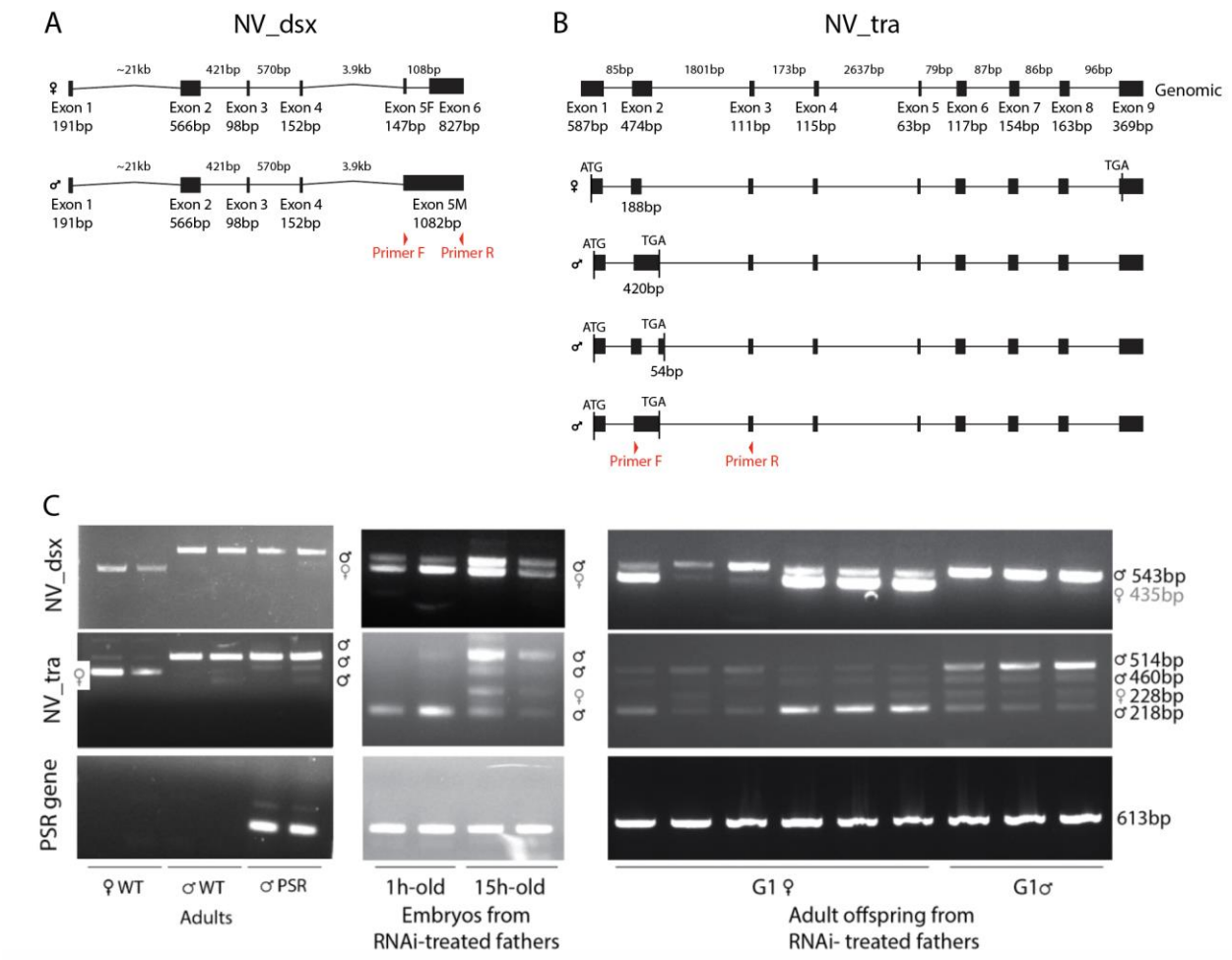


Fig.S11. *dsx* and *tra* splicing

A) Schematic representation of *Nasonia doublesex* female and male transcripts (NV_*dsx*). Exons are indicated with boxes and intron with lines. Red arrows indicate the position of primers for RT-PCR. B) schematic representation of *Nasonia transformer* gene and male and female specific transcripts (NV_tra). Exons are indicated with boxes and intron with lines. Red arrows indicate the position of primers for RT-PCR. **C)** Gel electrophoresis from RT-PCR depicting differential splicing variants of *NV_dsx* (*doublesex*), *NV_tra* (*transformer*), and PSR specific gene in wild type females and males (WT), PSR males, 1h and 15h-old embryos from RNAi-treated fathers and G1 females and males originated from RNAi-treated fathers. *NV_dsx* and *NV_tra* diagnostic primers from *(27)*. PSR gene analysed using PSR specific primers (Table S10).

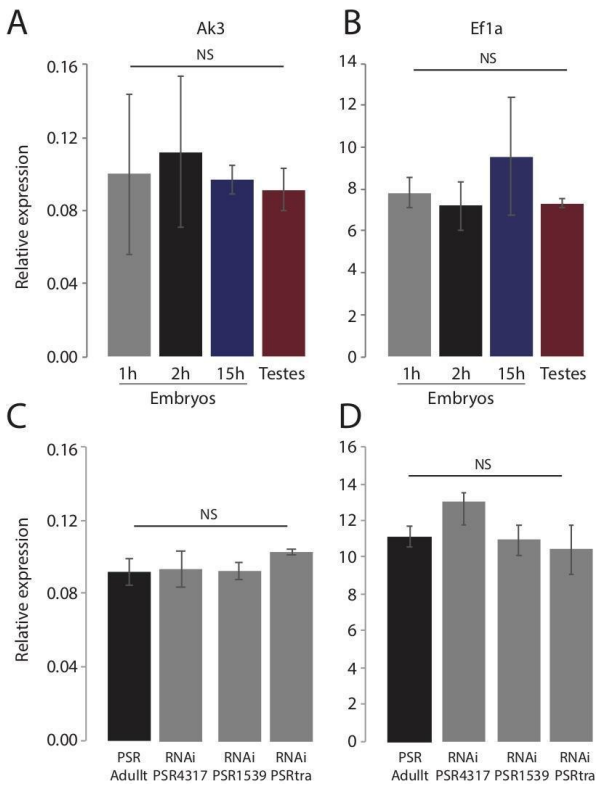


Fig.S12. Stability of reference genes between samples and treatments

The average relative expression of **A)** Ak3 and **B)** EF1α is compared among time points and **C)**, **D)** among RNAi- treated samples by two-way ANOVA.

Additional File

Supplementary Tables

Table S1 to S10

Table S1. BUSCO scores

Genome completeness assessed using BUSCO pipeline (*21*).

Table S2. PSR-specific contigs identification

List of all contigs and their cq ratio (chromosome quotient (*20*)) used to identify PSR-specific contigs. Cq calculates the ratio of the number of wild type and PSR+ reads mapping to a contig.

Table S3. Assembly statistics

Information about assembly of wild type Nasonia vitripennis genome and PSR chromosome

Table S4. Placed contigs

Information about contig numbers and length

Table S5. Summary of PSR composition

Summary of sequences present on PSR and their abundance.

Table S6. Repeats family and abundance on PSR

List of PSR repeats family, location and type of repeats.

Table S7. PSR specific transcripts

List of PSR expressed transcripts from Nanopore RNA sequencing of PSR-carrying testes and whole animal.

Table S8. PSR specific genes

List of PSR genes with location on PSR scaffold, Blast results, Pfam domain and expression data (TPM). TPM expression value are calculated with featureCounts (*27*) using Nanopore RNAseq data and an additional illumina RNA seq dataset from (*18*).

Table S9. Sex Ratio of G1 females and males after RNAi

Sex ratio of individual G1 females and males and presence of PSR

Table S10. Primers used and application

List of primers used for PCR, qPCR and dsRNA production.

Data S1-S6

Data S1. Genome assembly files

The archive contains a file with all the contigs in the assembly (Nvit_psr_1.fsa), the PSR specific contigs (Nvit_psr_1.psr_specific.fsa), the positions of placed contigs within chromosomes (Nvit_psr_1.chromosomes.agp) and chromosomal scaffolds generated using genetic markers (*23*).

Data S2. NCBI gene models

NCBI gene models have been mapped to our current assembly

Data S3. OGS2 gene models

OGS2 gene models have been mapped to our current assembly

Data S4. PSR most expressed transcript

Fasta file including the sequence of the three highest expressed PSR transcripts: haploidizer (PSR4317), PSR1539 and PSRtra.

Data S5. Blast results for *haploidizer*

Bastx results for haploidizer containing the top 20 hits and alignments.

Data S6. Gene prediction

The GTF file includes gene models supported by at least 10 full length aligned nanopore reads generated with Pinfish pipeline (Oxford Nanopore Technologies).