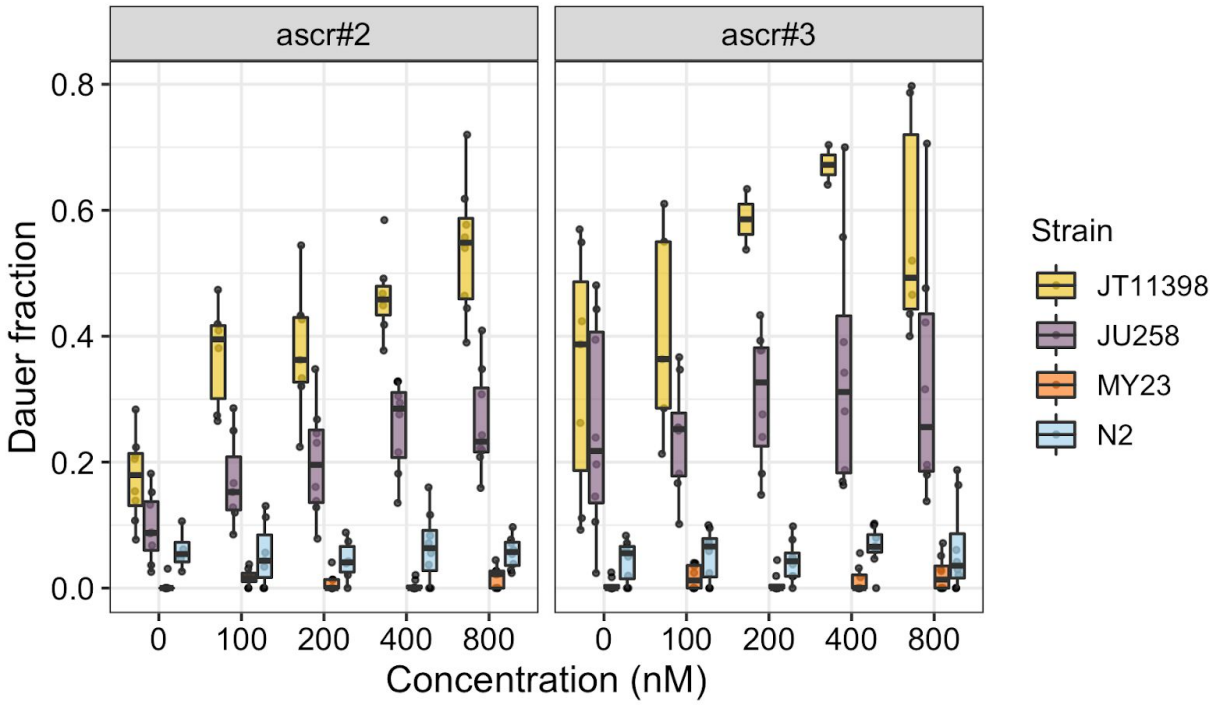
Supplementary Information for

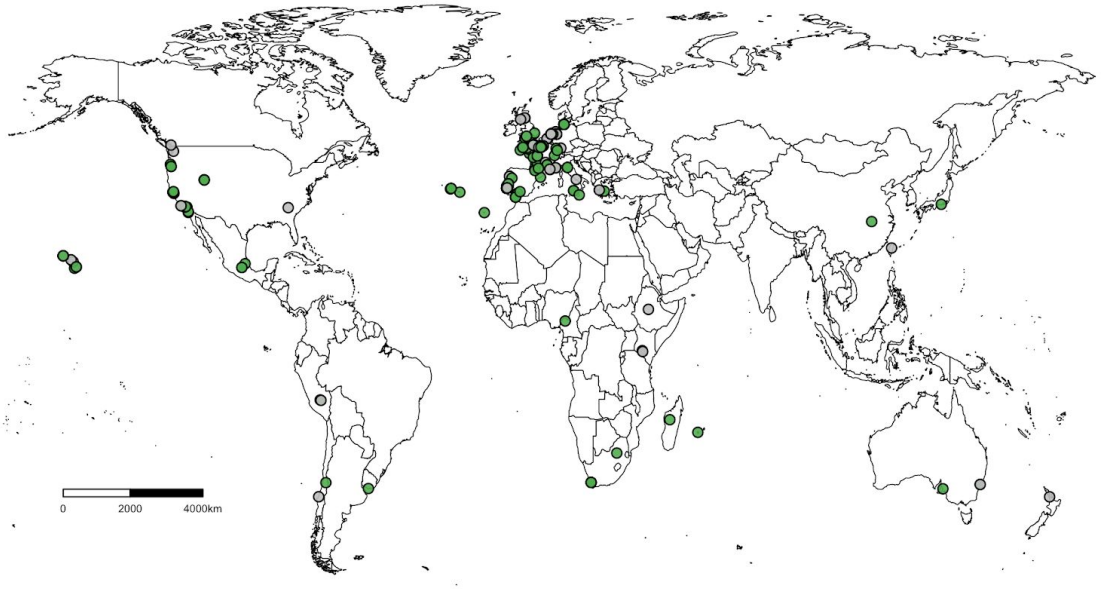
Selection and gene flow shape niche-associated variation in pheromone response

Daehan Lee, Stefan Zdraljevic, Daniel E. Cook, Lise Frézal, Jung-Chen Hsu, Mark G. Sterken, Joost A.G. Riksen, John Wang, Jan E. Kammenga, Christian Braendle, Marie-Anne Félix, Frank C. Schroeder, Erik C. Andersen



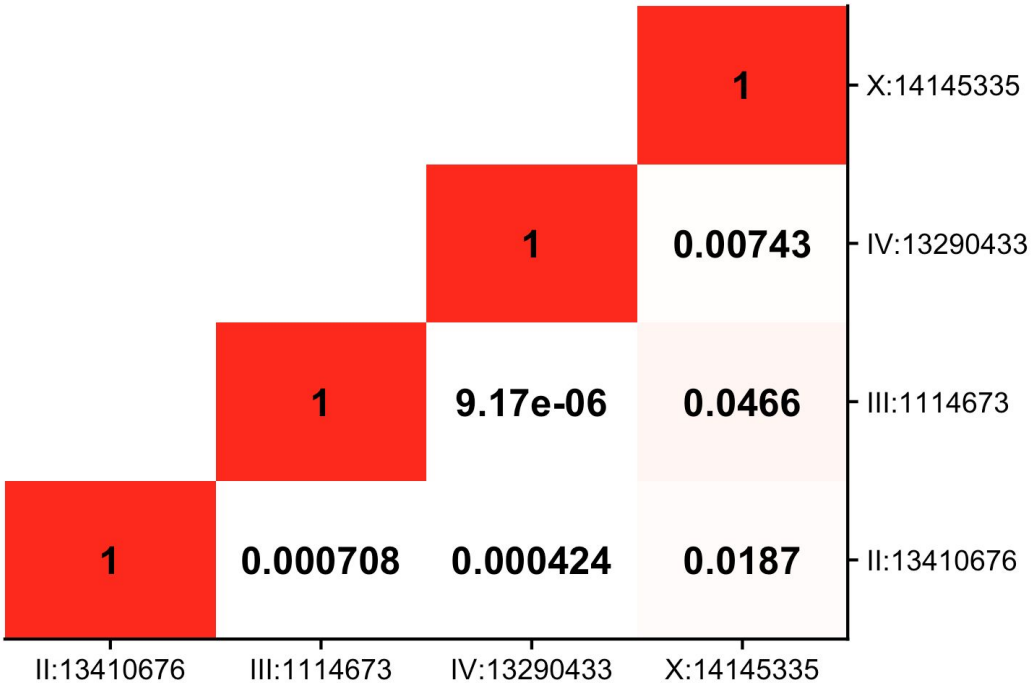
Supplementary Fig. 1: Genetically divergent wild isolates show different dose responses to dauer pheromone (ascr#2 and ascr#3) treatments

Tukey box plots of the dose responses at 25ºC for ascr#2 (left) and ascr#3 (right) for four divergent strains are shown with data points plotted behind. Box plots are colored by strain (JT11398 (yellow), JU258 (purple), MY23 (orange) and N2 (skyblue)). Concentrations of ascarosides are shown on the x-axis, and fractions of dauer larvae are shown on the y-axis. The horizontal line in the middle of the box is the median, and the box denotes the 25th to 75th quantiles of the data. The vertical line represents the 1.5 interquartile range.



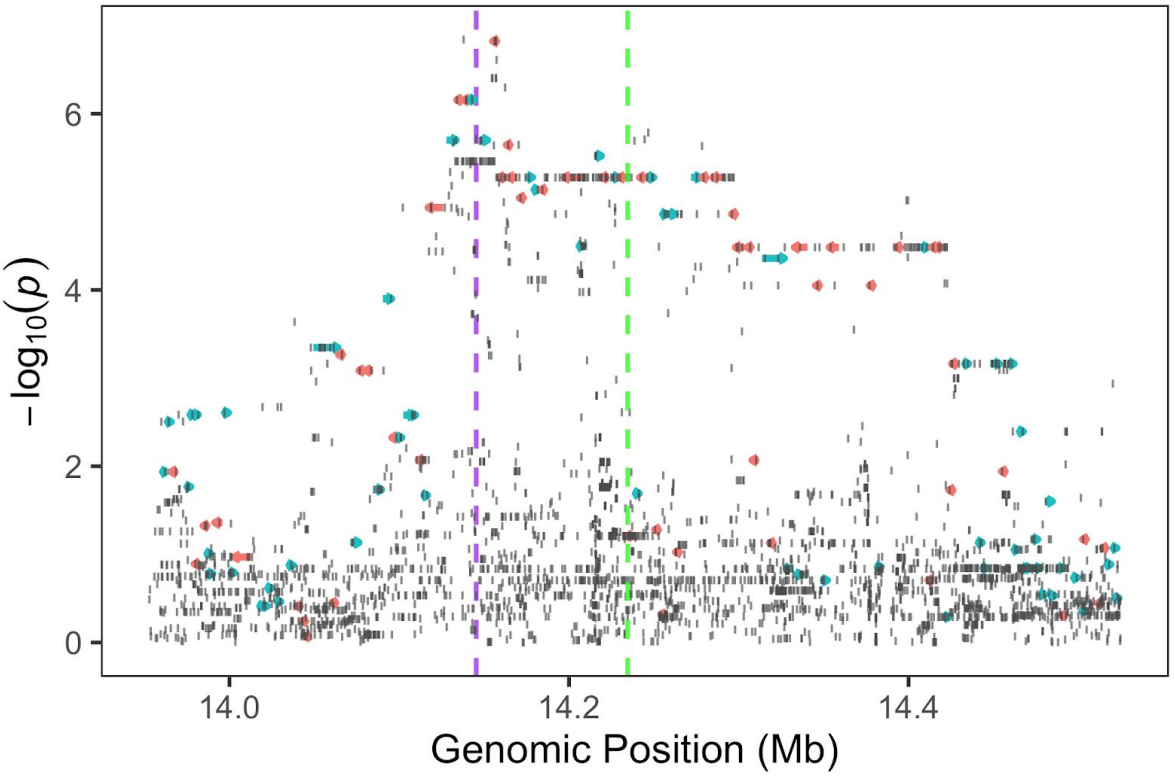
Supplementary Fig. 2: Wild *C. elegans* strains have isolated across six continents

Among the 249 wild strains that are available through the CeNDR, the global distribution of the 239 wild strains with known geographic origins is shown. Among the 157 wild strains that were tested using the HTDA, 151 wild strains with known geographic origins are shown as green circles. Wild strains that are not tested using the HTDA are shown as grey circles. A scale bar is shown in the map.



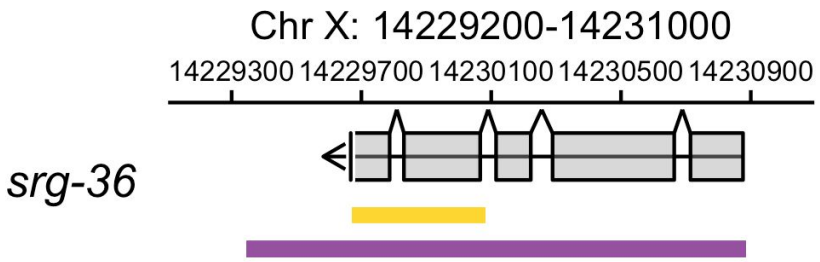
Supplementary Fig. 3: Linkage disequilibrium is not observed among four QTL for ascr#5 response

A heatmap plot that shows linkage disequilibrium (LD) among peak QTL markers as measured by the square of the correlation coefficient (r^2).



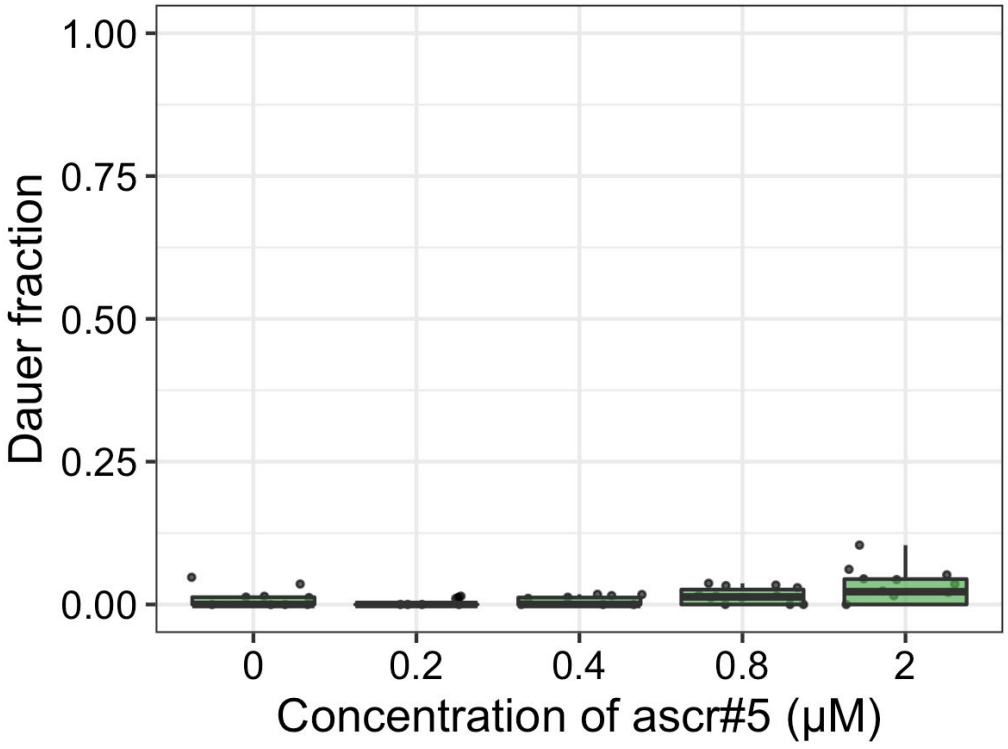
Supplementary Fig. 4: *dauf-1* QTL contains two ascr#5 receptor genes, *srg-36* and *srg-37*

A plot for fine-mapping of ascr#5 response across the *dauf-1* QTL. Each grey bar represents a SNV or indel that is present in at least 5% of the 157 wild strains assayed for dauer formation. The genomic position in Mb is plotted on the x-axis, and the statistical significance of the correlation between genotype and phenotype is plotted on the y-axis. Genes in the region are labeled as colored arrows (skyblue: forward, orange: reverse). The peak of *dauf-1* QTL from GWA mapping is represented by vertical purple line. The location of *srg-36* and *srg-37* is represented by vertical green line.



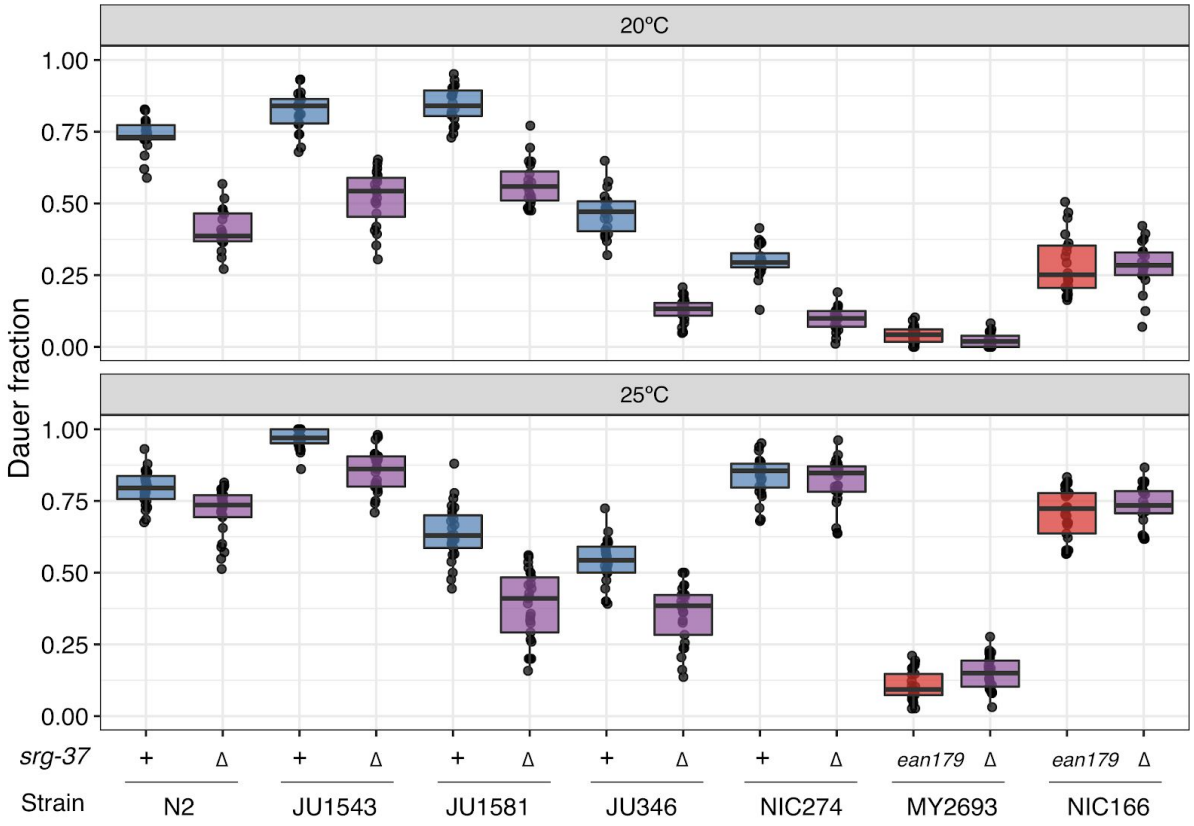
Supplementary Fig. 5: A schematic plot for *srg-36* gene structure

A schematic plot for the *srg-36* gene structure (grey), the 411-bp natural deletion allele *ean178* (yellow), and the CRISPR-Cas9 loss-of-function deletion (purple) are shown.



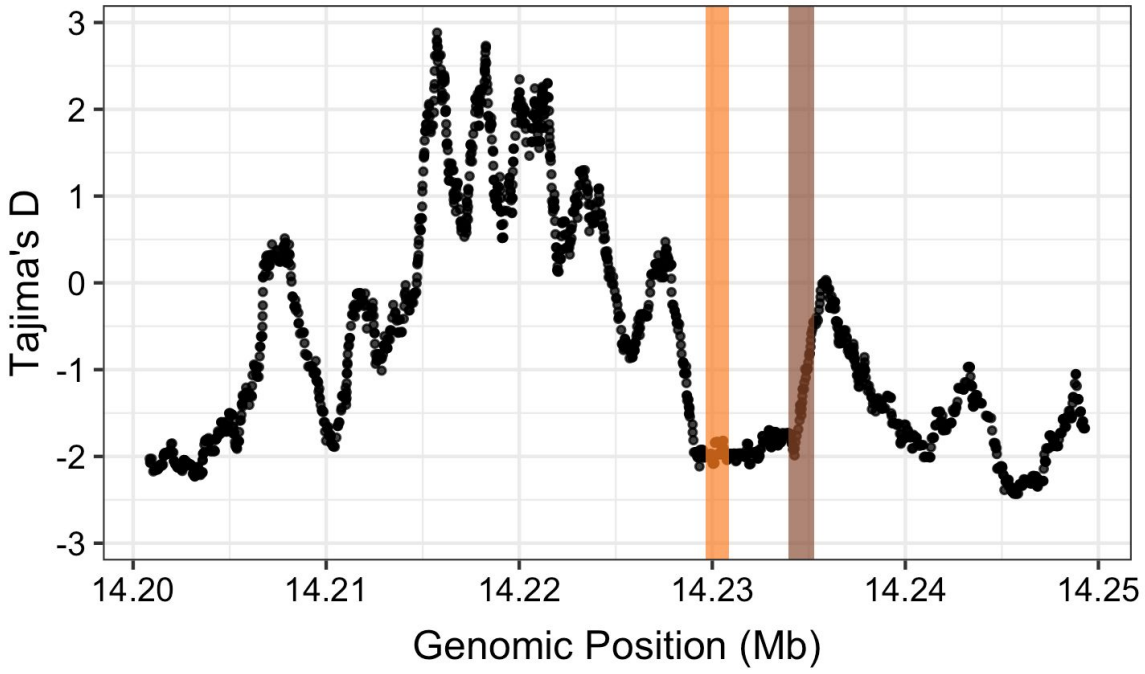
Supplementary Fig. 6: PB303, a unique strain that has a deleted *srg-36*, is insensitive to ascr#5

Tukey box plots of the ascr#5 dose response at 25ºC for PB303 that carries the *srg-36(ean178)* deletion allele are shown with data points plotted behind. Concentrations of ascr#5 are shown on the x-axis, and the fraction of dauer formation is shown on the y-axis. The horizontal line in the middle of the box is the median, and the box denotes the 25th to 75th quantiles of the data. The vertical line represents the 1.5 interquartile range.



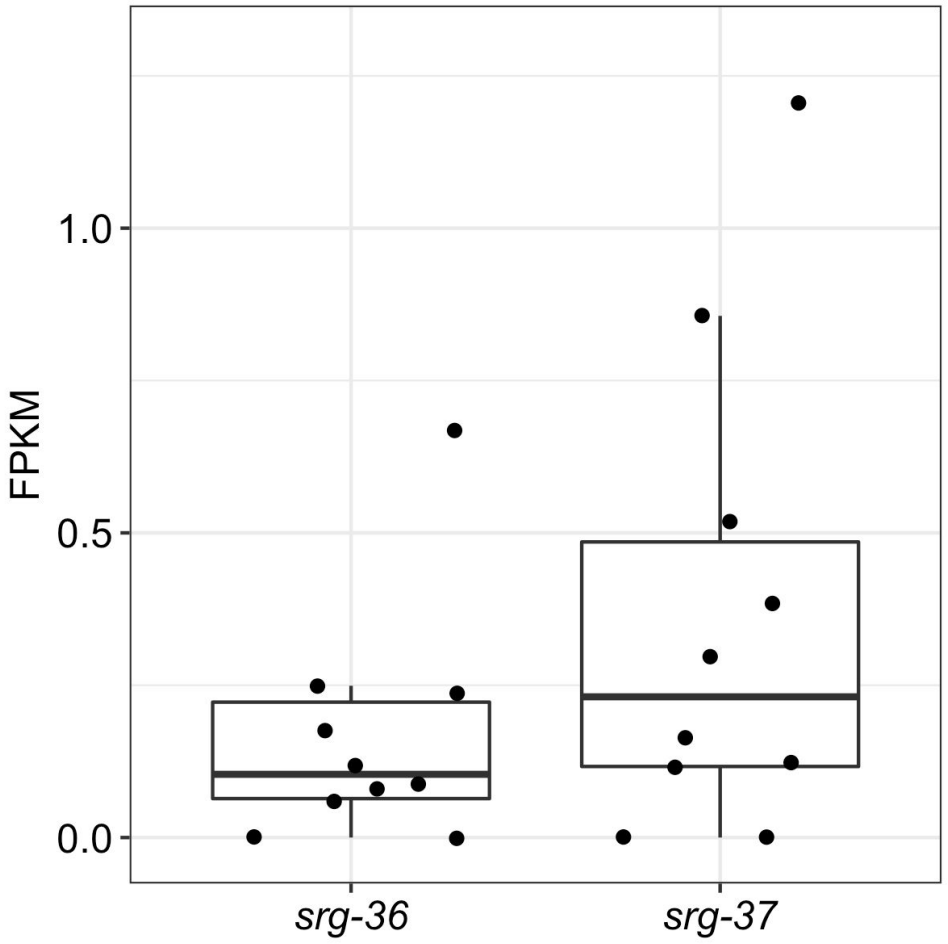
Supplementary Fig. 7: The *srg-37(ean179)* deletion is a putative loss-of-function allele

Tukey box plots of *srg-37* loss-of-function experiments in dauer pheromone conditions (800 nM of ascr#5) at 20ºC (top) and 25ºC (bottom) are shown with data points plotted behind. Each box plot is colored by the genotypes of *srg-37*, where *srg-37(+)* is blue, *srg-37(ean179)* is red, and CRISPR-Cas9-generated *srg-37(lf)* is purple. Wild isolates and the putative *srg-37* loss-of-function deletion mutants for each genetic background (Δ) are paired on the x-axis. The fraction of dauer formation is shown on the y-axis. The horizontal line in the middle of the box is the median, and the box denotes the 25th to 75th quantiles of the data. The vertical line represents the 1.5 interquartile range.



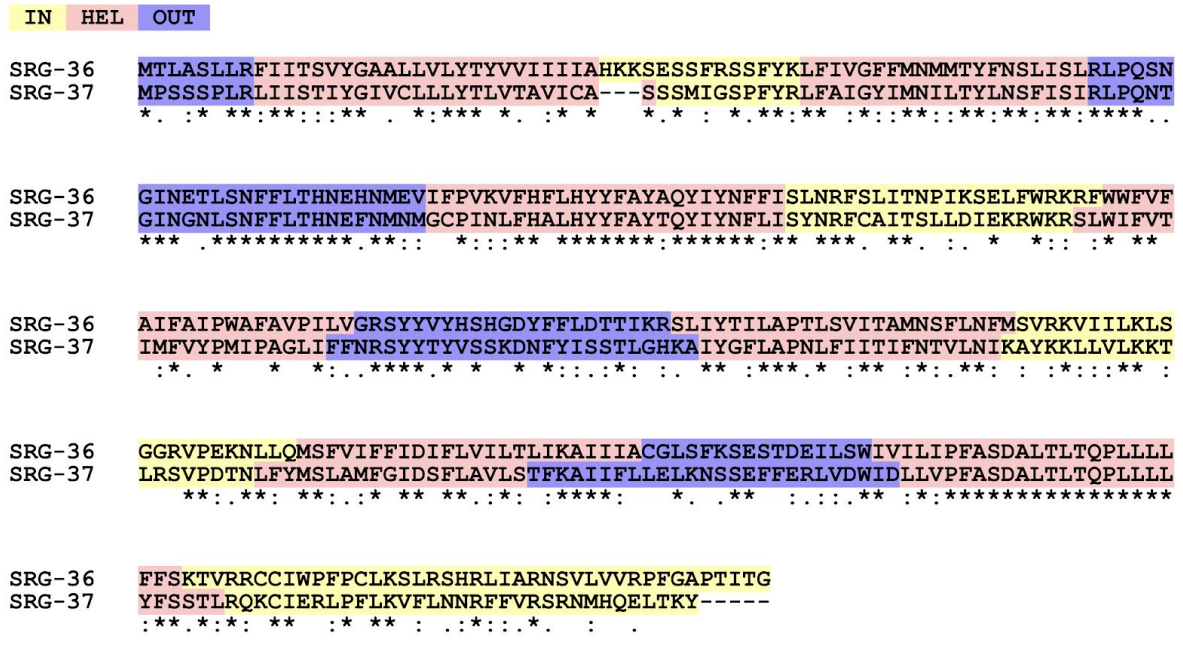
Supplementary Fig. 8: Tajima's D statistics decrease around *srg-36* gene

Tajima's D statistics for a region of X chromosome including *srg-36 srg-37* locus are shown. Each dot corresponds to a Tajima's D statistic calculated from the allele frequency spectrum of 50 SNVs across 249 wild isolates. Colored lines highlight coding regions of *srg-36* (orange) and *srg-37* (brown).



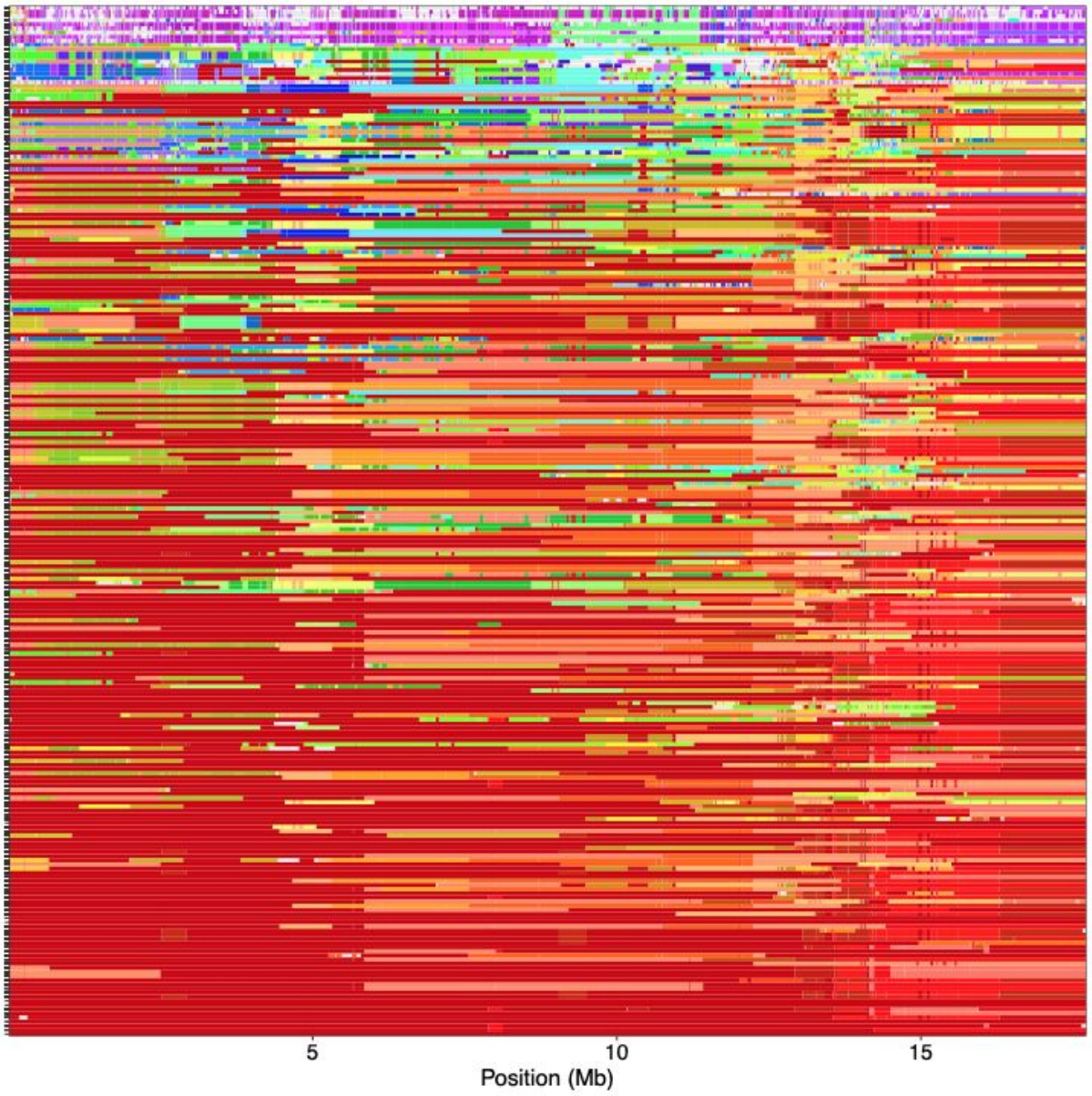
Supplementary Fig. 9: Gene expression levels are not significantly different between *srg-36* and *srg-37*

Tukey box plots of the gene expression levels of *srg-36* and *srg-37*. Each dot corresponds to average FPKM value of each gene at L1 stage from independent RNA-seq experiments. The horizontal line in the middle of the box is the median, and the box denotes the 25th to 75th quantiles of the data. The vertical line represents the 1.5 interquartile range.



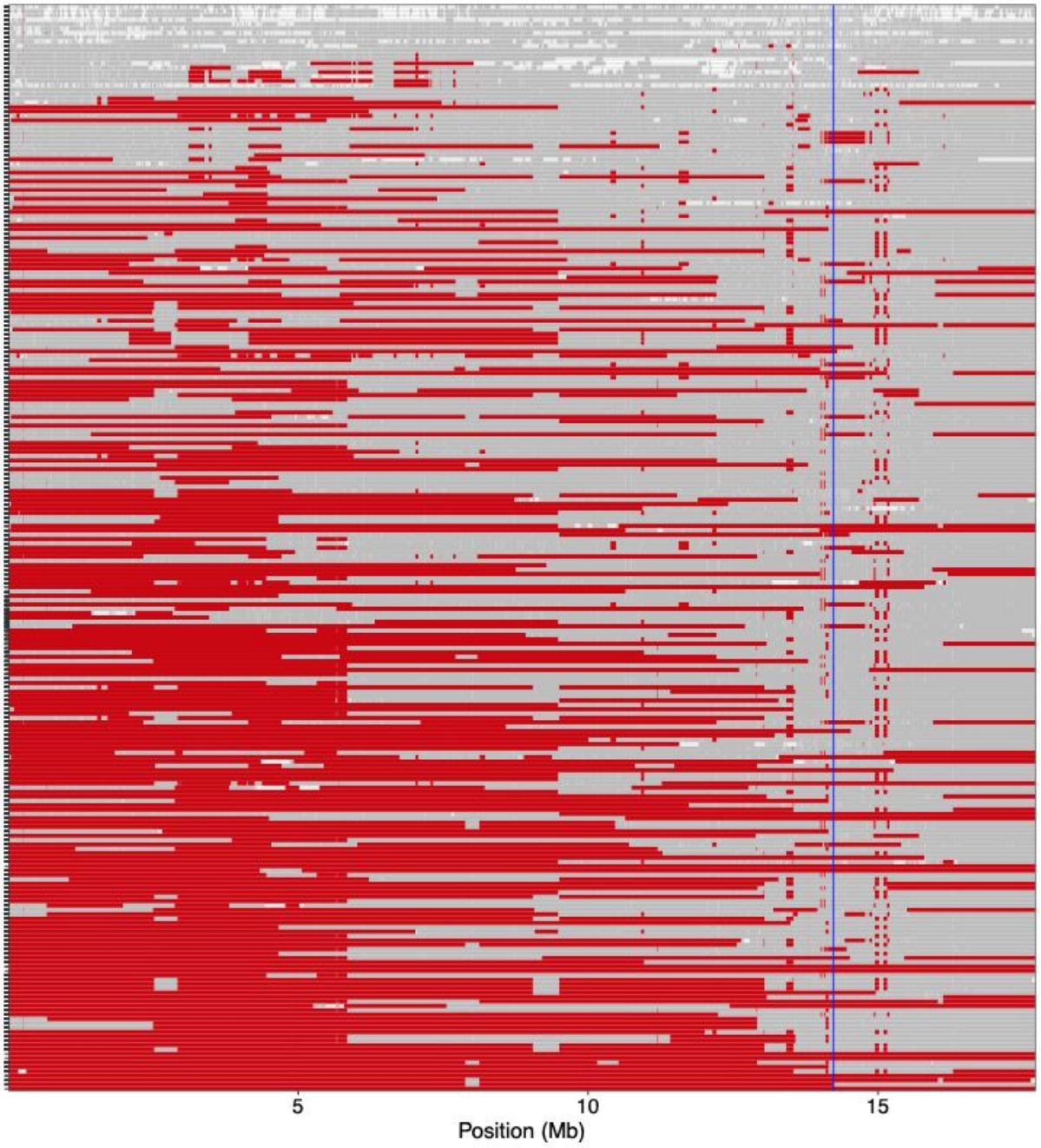
Supplementary Fig. 10: An alignment of SRG-36 and SRG-37 shows conserved transmembrane receptor structure

Reference sequences from the N2 strain were used. Alignment and structure estimations were performed by PSI-TM/Coffee software. Amino acid residues of each protein are colored by putative structural domains (intracellular domains (yellow), transmembrane domain (pink), extracellular domains (blue)). Conservation is annotated below each aligned amino acid (conserved residue (asterisk), strongly conserved residue (> 0.5 in the Gonnet PAM 250 matrix, colon), weakly conserved residues (=< 0.5 in the Gonnet PAM 250 matrix, period).



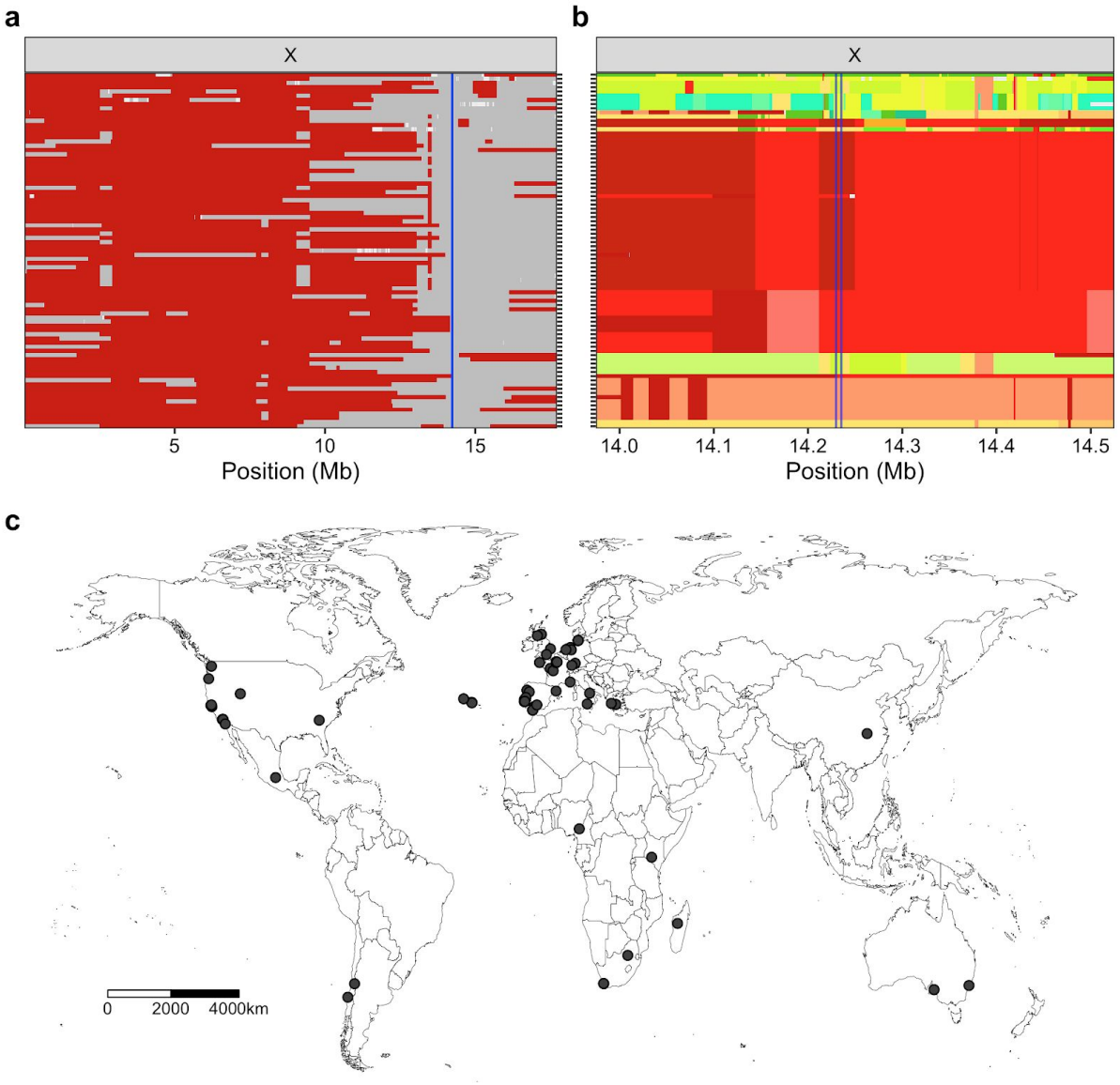
Supplementary Fig. 11: X chromosome sharing of 249 wild strains reveals swept regions where haplotype diversity is reduced

X chromosome sharing plots with color-coded haplotype segments of the 249 wild strains of *C. elegans* are shown. Each row corresponds to genetically unique single genome-wide genotypes (isotype), ordered roughly by the extent of haplotype sharing. Shared chromosome regions are shown with the same color. Genomic position on the X chromosomes is shown on the x-axis. Note that similarity in colors does not imply similarity in genotype.



Supplementary Fig. 12: *srg-37* is located far from the left arm of the X chromosome, which is frequently swept across 249 wild *C. elegans* strains

A modified X chromosome sharing plots with the swept haplotype colored as red and other haplotypes as grey are shown. The swept haplotype is defined as a haplotype that is identical-by-descent and most frequently shared across 249 wild *C. elegans* strains. Each row shows the genotype of one of the 249 wild isolates, ordered by the extent of swept haplotype sharing. Genomic position is shown on the x-axis. The blue vertical lines denote the *srg-36 srg-37* locus.


a
X
5
10
15
Position (Mb)
b
X
14.0
14.1
14.2
14.3
14.4
14.5
Position (Mb)
c
0
2000
4000km

Supplementary Fig. 13: *srg-37(ean179)* is outcrossed by multiple genotypes

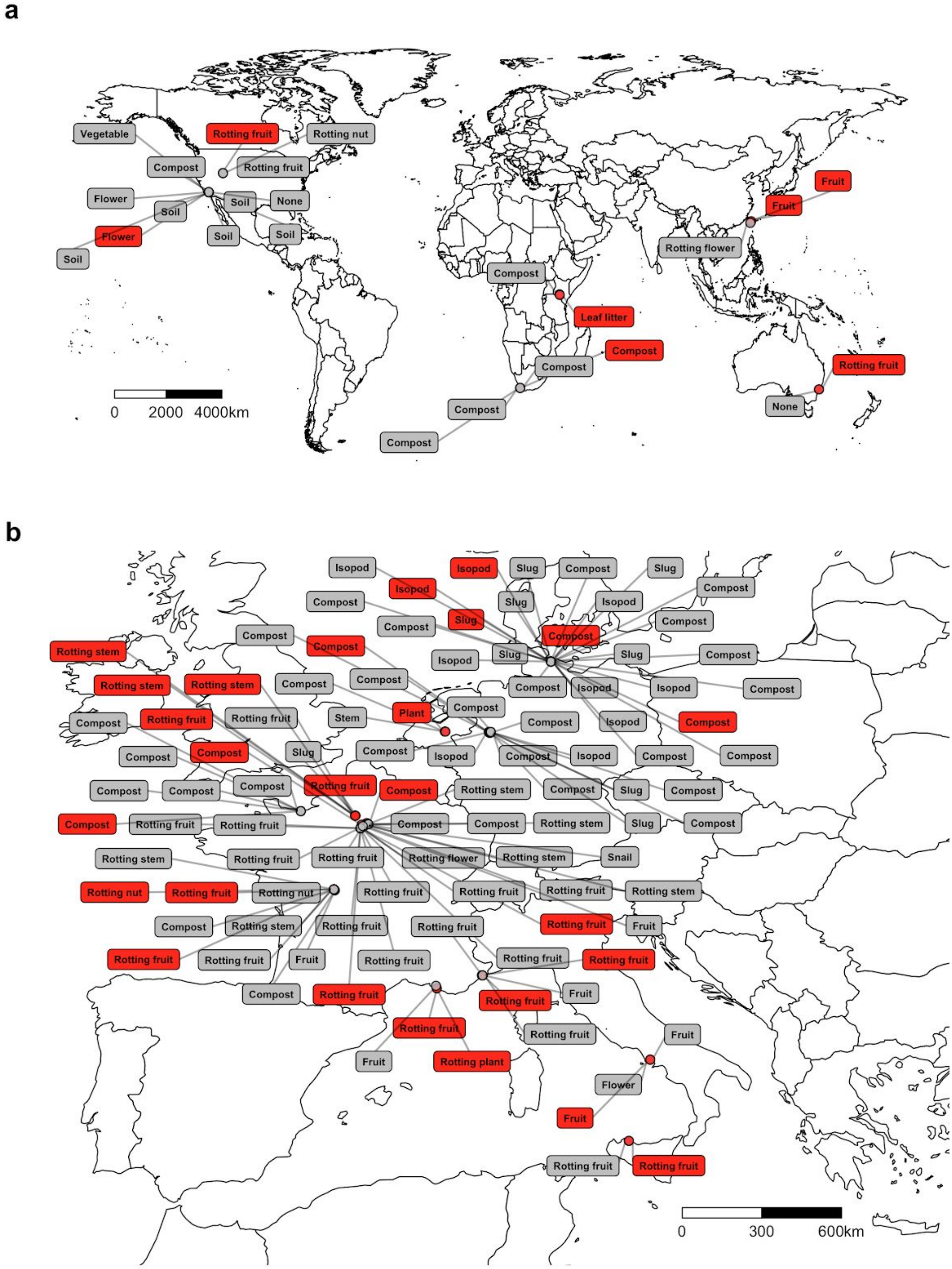
The outcrossed subpopulation is defined as a subset of strains with the swept haplotype segments that span more than 50% of the X chromosome except *srg-37* locus. This outcrossed subpopulation comprises 85 wild strains.

(a) A modified haplotype sharing plot with the swept haplotype colored as red and other haplotypes colored grey. Each row is one of the wild isolates that belongs to the outcrossed subpopulation, ordered roughly by the extent of the swept haplotype sharing.

(b) A haplotype sharing plot with color-coded haplotype segments of outcrossed subpopulation across the X chromosome region (X: 14.0 Mb - 14.5 Mb) flanking and including *srg-36* and *srg-37*. Note that red and dark red color are not swept haplotypes. All shown haplotypes were color-coded by grey in (a).

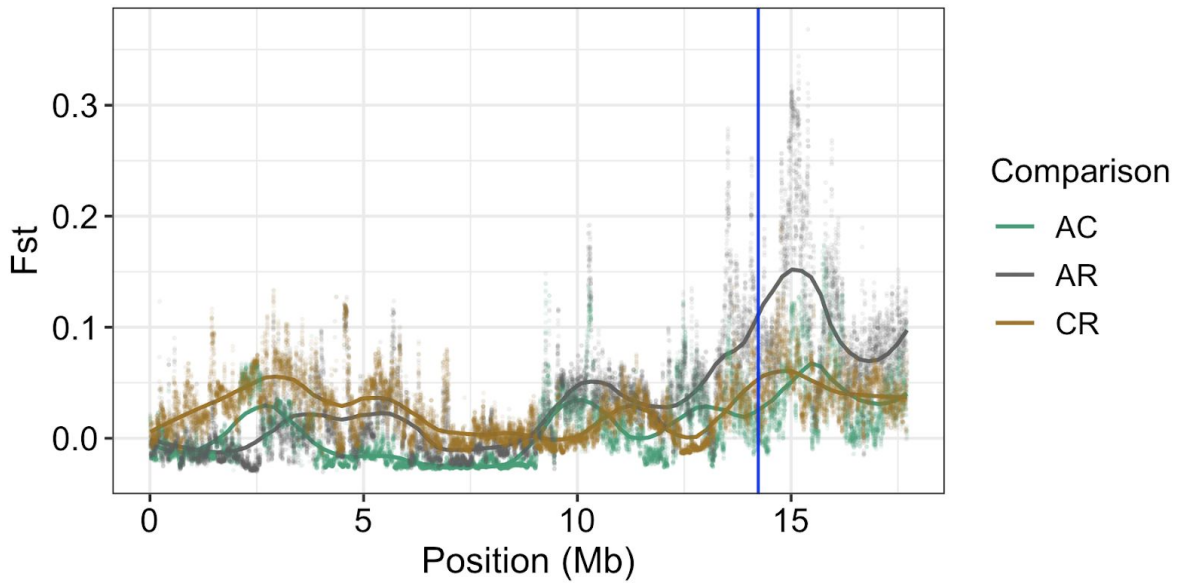
(a, b) The blue vertical lines denote the *srg-36 srg-37* locus.

(c) The geographic locations of 85 outcrossed wild strains are shown as grey circles. A scale bar is shown in the map.


a
Vegetable
Rotting fruit
Rotting nut
Compost
Rotting fruit
Flower
Soil
None
Soil
Flower
Soil
Soil
Soil
Fruit
Fruit
Rotting flower
Compost
Leaf litter
Compost
Compost
Rotting fruit
None
Compost
Compost
0 2000 4000km
b
Isopod
Isopod
Slug
Compost
Slug
Isopod
Compost
Slug
Isopod
Compost
Compost
Slug
Compost
Compost
Compost
Compost
Rotting stem
Isopod
Slug
Slug
Compost
Rotting stem
Rotting stem
Compost
Compost
Compost
Isopod
Isopod
Compost
Compost
Rotting fruit
Rotting fruit
Stem
Plant
Compost
Compost
Isopod
Compost
Compost
Compost
Slug
Compost
Isopod
Compost
Isopod
Compost
Compost
Compost
Compost
Rotting fruit
Compost
Rotting stem
Compost
Slug
Compost
Compost
Rotting fruit
Rotting fruit
Compost
Compost
Rotting stem
Slug
Compost
Rotting stem
Rotting fruit
Rotting fruit
Rotting flower
Rotting stem
Snail
Rotting nut
Rotting fruit
Rotting nut
Rotting fruit
Rotting fruit
Rotting fruit
Rotting stem
Compost
Rotting stem
Rotting fruit
Rotting fruit
Rotting fruit
Fruit
Rotting fruit
Rotting fruit
Fruit
Rotting fruit
Rotting fruit
Rotting fruit
Compost
Rotting fruit
Rotting fruit
Fruit
Rotting fruit
Rotting fruit
Fruit
Rotting plant
Fruit
Flower
Fruit
Rotting fruit
Rotting fruit
0 300 600km

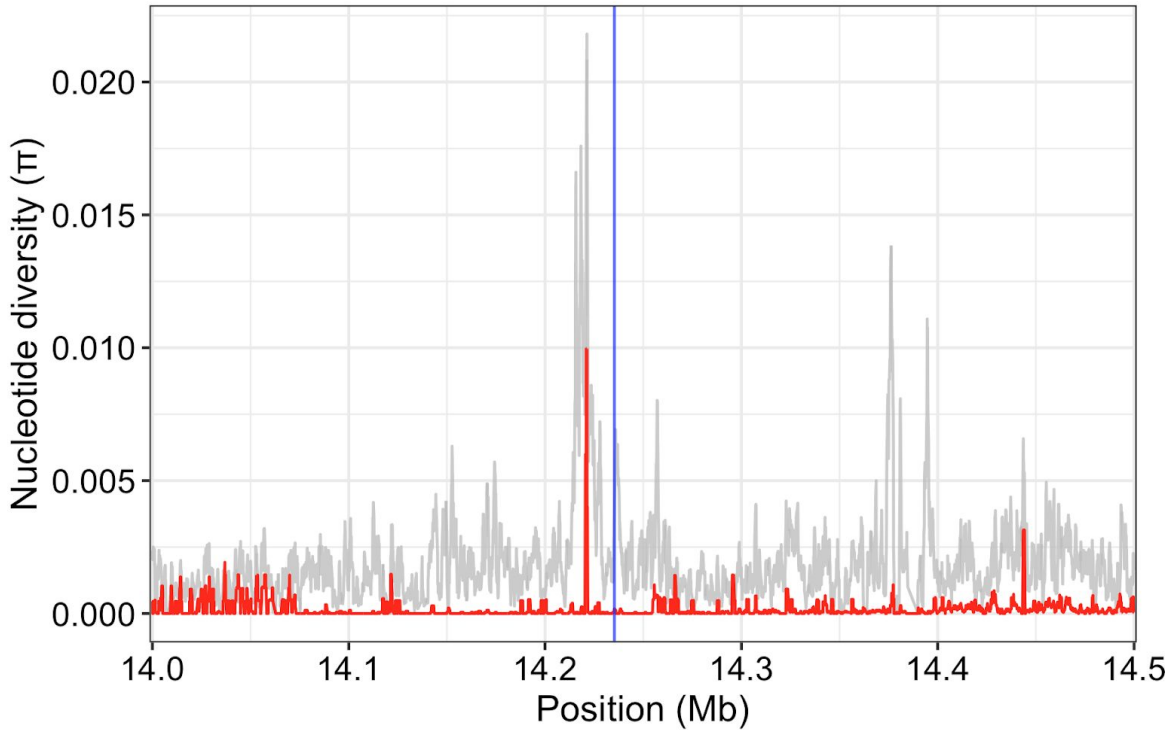
Supplementary Fig. 14: Two *srg-37* genotypes are isolated in the local habitat across the world

Wild strains that are isolated from locations where both genotypes of *srg-37* were found are shown on the map of (a) world and (b) Europe. Each box indicates the substrate type where each strain is isolated, and is colored by the *srg-37* genotype of the isolated strain (grey: wild-type, red: *ean179*). Scale bars are shown in the map.



Supplementary Fig. 15: High genetic divergence at *srg-37* locus is observed between niche-associated subpopulations

The X chromosome-wide F_{ST} statistics between animal-compost (AC, green), animal-rotting fruit (AR, grey), and compost-rotting fruit (CR, brown) subpopulations from Fig. 4e are shown. F_{ST} was calculated with a window size of 10 kb and a step size of 1 kb. Each dot corresponds to the F_{ST} statistic of an individual window. Smoothed lines are LOESS fits. The blue vertical line denotes the location of the *srg-37* gene.



Supplementary Fig. 16: Nucleotide diversity (π) across *srg-37* locus is low for wild strains with the *srg-37* deletion

Line plots for nucleotide diversity (π) across the *srg-37* locus is shown. π was calculated for the subpopulation with the *srg-37(+)* allele (grey) and the subpopulation with the *srg-37(ean179)* deletion allele (red) with a window size of 1 kb and a step size of 100 bp. The blue vertical line denotes the location of the *srg-37* gene.

Supplementary Table 1: Oligonucleotides and crRNAs used in this study

Class	ID	Sequences	Description	Source
Oligonucleotide	oECA1429	GTAACCGCTGTCATCTGTGC	*srg-37* genotyping external/forward primer	IDT
Oligonucleotide	oECA1430	CCTGATGCATATTTCTTGATCTG	*srg-37* genotyping reverse primer	IDT
Oligonucleotide	oECA1435	TGGATTCCTTGCTCCAAATC	*srg-37* genotyping internal/forward primer	IDT
Oligonucleotide	oECA1460	CTTTGAGAATATGACACGCCC	*srg-36* genotyping external/forward primer	IDT
Oligonucleotide	oECA1461	CTCGACACGCTGTGTTTCTA	*srg-36* genotyping external/reverse primer	IDT
Oligonucleotide	oECA1462	CAACGCTGTCCGTGATAACT	*srg-36* genotyping internal/forward primer	IDT
Oligonucleotide	oECA1463	CTCGACACGCTGTGTTTCTA	*srg-36* genotyping internal/reverse primer	IDT
crRNA	crECA76	TAAAGCTGTTTAGATAGGTA	*srg-37* exon 2 targeting crRNA	IDT
crRNA	crECA77	CGAACGTTTACCCTTTCTCA	*srg-37* exon 5 targeting crRNA	IDT
crRNA	crECA82	GCCAGCGTCATACCAACAGG	*srg-36* 5’ targeting crRNA	IDT
crRNA	crECA83	TAACGCGAATAGAGGATCGG	*srg-36* 3’ targeting crRNA	IDT