bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Selection and gene flow shape niche-associated copy-number variation of pheromone receptor genes

Daehan Lee[1], Stefan Zdraljevic[1,2], Daniel E. Cook[1,2,3] , Lise Frézal[4], Jung-Chen Hsu[5], Mark G. Sterken[6], Joost A.G. Riksen[6], John Wang[5], Jan E. Kammenga[6], Christian Braendle[7], Marie-Anne Félix[4], Frank C. Schroeder[8], Erik C. Andersen[1]*

1. Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208, USA
2. Interdisciplinary Biological Sciences Program, Northwestern University, Evanston, IL 60208, USA
3. Present address: The Francis Crick Institute, London NW1 1ST, UK
4. Institut de Biologie de l'Ecole Normale Supérieure, Centre National de la Recherche Scientifique, INSERM, École Normale Supérieure, Paris Sciences et Lettres, Paris, France
5. Biodiversity Research Center, Academia Sinica, Taipei, 11529, Taiwan
6. Laboratory of Nematology, Wageningen University and Research, 6708PB, The Netherlands
7. Université Côte d'Azur, CNRS, Inserm, IBV, France, 06100 Nice, France
8. Boyce Thompson Institute and Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853, USA

* Corresponding author

Erik C. Andersen
Assistant Professor of Molecular Biosciences
Northwestern University
Evanston, IL 60208, USA
Tel: (847) 467-4382
Fax: (847) 491-4461
Email: Erik.Andersen@Northwestern.edu

Daehan Lee, daehan.lee@northwestern.edu, ORCID 0000-0002-0546-8484
Stefan Zdraljevic, stefanzdraljevic2018@u.northwestern.edu, ORCID 0000-0003-2883-4616
Daniel E. Cook, danielecook@gmail.com, ORCID 0000-0003-3347-562X
Lise Frézal, frezal@biologie.ens.fr
Jung-Chen Hsu, ronjan0923@gmail.com
Mark G. Sterken, mark.sterken@wur.nl, ORCID 0000-0001-7119-6213
Joost A.G. Riksen, joost.riksen@wur.nl
John Wang, johnwang@gate.sinica.edu.tw, ORCID 0000-0002-6179-775X
Jan E. Kammenga, jan.kammenga@wur.nl, ORCID 0000-0003-4822-4436
Christian Braendle, braendle@unice.fr, ORCID 0000-0003-0203-4581
Marie-Anne Félix, felix@biologie.ens.fr
Frank C. Schroeder, fs31@cornell.edu, ORCID 0000-0002-4420-0237
Erik Andersen, erik.andersen@northwestern.edu, ORCID 0000-0003-0229-9651

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Abstract

From quorum sensing in bacteria to pheromone signaling in social insects, chemical communication mediates interactions among individuals in a local population. In *Caenorhabditis elegans*, ascaroside pheromones can dictate local population density, in which high levels of pheromones inhibit the reproductive maturation of individuals. Little is known about how genetic diversity in natural population affects the pheromone responses of individuals from diverse habitats. Here, we show that a niche-associated copy-number variation (CNV) of pheromone receptor genes contributes to natural differences in pheromone responses. We found putative loss-of-function deletions that reduce copy number of two duplicated pheromone receptor genes (*srg-36 and srg-37*), which were shown previously to be selected in population-dense laboratory cultures. A common natural deletion in the less functional copy (*srg-37*) arose from a single ancestral population that spread throughout the world and underlies reduced pheromone sensitivity across the global *C. elegans* population. This deletion is enriched in wild strains that were isolated from a rotting fruit niche, where proliferating populations are often found. Taken together, these results demonstrate that selection and gene flow together shape the copy number of pheromone receptor genes in the natural *C. elegans* population to facilitate local adaptation in diverse niches.

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Introduction

To maximize reproductive success, organisms must respond to changing environmental conditions. In a fluctuating environment, each response will likely have a fitness trade-off with reproductive success now or in the future. *Caenorhabditis elegans* can either grow to a reproductive adult in three days or delay maturity for months by entering the dauer diapause stage[1]. Food supply and pheromone signal are two major inputs that affect this developmental plasticity[2]. The accumulation of ascaroside pheromones in dense populations antagonizes food signals that promote growth and inhibits the induction of the stress-resistant and long-lived dauer stage[3,4]. Animals must measure the amount of remaining food and the pheromone concentration to determine if it is advantageous to continue reproductive growth or enter the dauer stage to disperse and hopefully encounter a new food source. Therefore, the decision to enter dauer decreases reproductive success in the short-term in favor of future survival success. Decades of research have provided insights into the chemical and genetic basis of the dauer-pheromone response[5]. However, these studies used a single laboratory-adapted strain (N2), which has limited our understanding of the natural processes that have shaped the dauer-pheromone response.

After decades of focused laboratory research on *C. elegans* as a model organism, the natural history of this species has only recently been described from extensive field research[6]. These field studies have revealed that the dauer stage is important for the population dynamics in their natural habitat[7]. These dynamics are typified by a “boom” phase after initial colonization of a nutrient-rich habitat, followed by a “bust” phase when resources are depleted. At the end of the boom phase when the local population size is large and nutrients are limited, individual animals enter the dauer stage. Dauers exhibit a stage-specific behavior called nictation, which facilitates phoretic interactions between dauer larvae and more mobile animals to disperse to favorable environments[8,9]. Because dauer larvae are presumed to play a crucial role in the survival and dispersal of the species, it is

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

likely that the decision to enter the dauer stage is under natural selection. Here, we integrate laboratory experiments, computational genomic analyses, and field research to further our understanding of the genetic basis underlying microevolution of the pheromone-mediated developmental plasticity. We identify the natural genetic variation of dauer-pheromone responses and characterize a copy-number variation of pheromone receptor genes that has been shaped by niche-associated selection and gene flow.

RESULTS

Natural variation of the dauer-pheromone response was measured using a high-throughput dauer assay

To explore the effects of natural genetic variation on the ability to enter the dauer stage, we developed a high-throughput assay (HTDA) to quantify the dauer-pheromone responses of wild *C. elegans* strains. The HTDA takes advantage of the observation that dauer larvae have no pharyngeal pumping[10]. We treated animals with fluorescent microspheres that can be ingested and then quantified both fluorescence and size of individual animals using a large-particle flow cytometer (COPAS BIOSORT, Union Biometrica). These data facilitated computational classification of dauers (Fig. 1a, b; Materials and methods) and recapitulated the known differences in the dauer-pheromone responses between N2 and a constitutive dauer mutant *daf-2(e1370)*, as well as the dauer-inducing effect of synthetic pheromone (Fig. 1b, c). To determine if genetic variation within the *C. elegans* species causes differential dauer-pheromone responses, we applied the HTDA to four genetically divergent *C. elegans* strains after treatment with various concentrations of four known dauer-inducing synthetic ascarosides (ascr#2, ascr#3, ascr#5, and ascr#8). We found significant variation in the dauer-pheromone response among the strains tested, as measured by the fraction of individuals that enter the dauer stage (Fig. 1d,

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Supplementary Fig. 1). Among the conditions we tested, we found that 800 nM ascr#5 maximizes the among-strain variance and minimizes the within-strain variance in dauer-pheromone response. These results enabled us to survey the effects of genetic variation on the dauer-pheromone response across the *C. elegans* species.

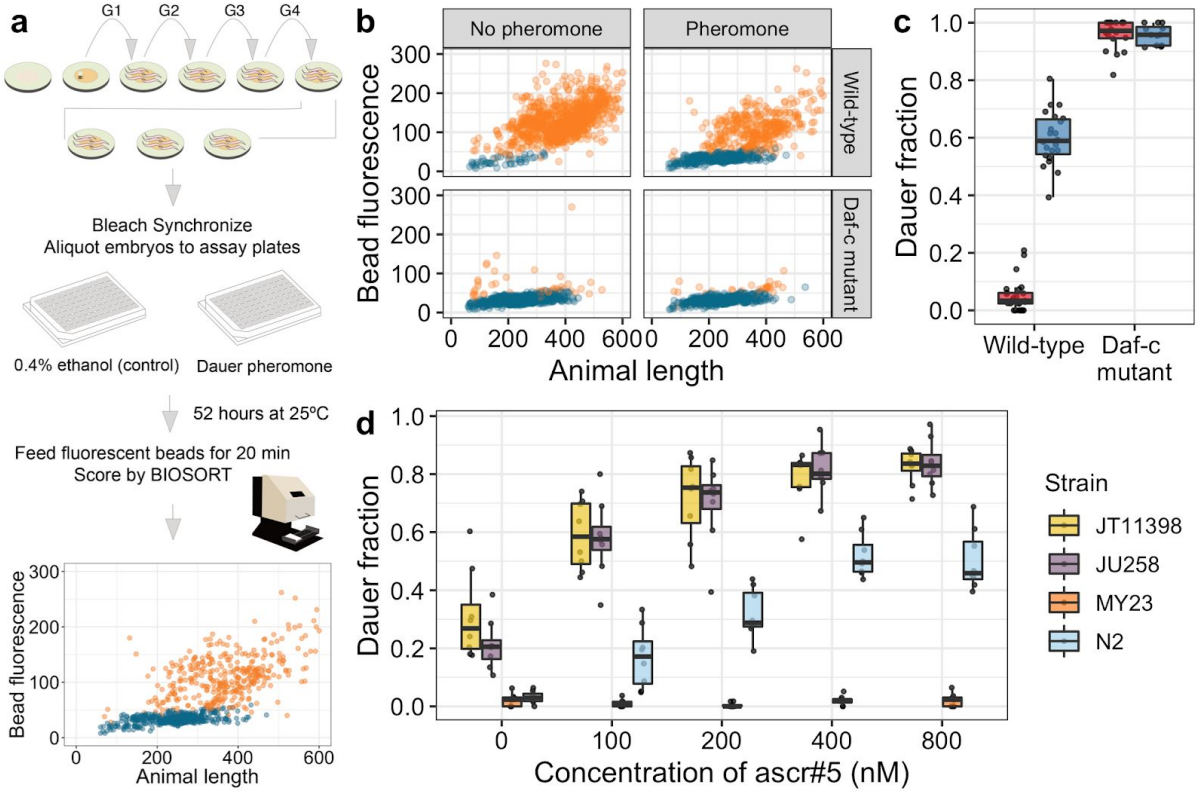

Fig. 1: A high-throughput dauer assay measures natural variation of dauer-pheromone response

(a) The workflow for the high-throughput dauer assay (HTDA) using a COPAS BIOSORT is shown (see Materials and methods for further description). (b) Optical measurements of the laboratory wild-type strain (N2) (Top) and a Daf-c mutant *daf-2(e1370)* (bottom) are shown under control (left) and pheromone-treated (ascr#5 800 nM) conditions (right) at 25ºC using the HTDA. Animal size and fluorescence-intensity traits are used as variables to build a model that differentiates dauer (blue) and non-dauer populations (orange). Relative animal length measured by time-of-flight (TOF) is shown on the x-axis, and bead-derived fluorescent intensity is shown on the y-axis. (c) Tukey box plots of the dauer fraction quantification from (b) are shown with data points plotted behind. Box plots are colored by assay conditions (control (red) and ascr#5 800 nM treatment (blue)). The genotypes are shown on the x-axis, and fractions of dauer larvae are shown on the y-axis. (d) Tukey box plots of the ascr#5 dose response at 25ºC 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 (blue)). Concentrations of ascr#5 are shown on the x-axis, and fractions of dauer larvae are shown on the y-axis.

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Genome-wide association (GWA) mapping reveals multiple loci underlying natural variation of the ascr#5 response

Next, we quantified dauer induction of 157 wild strains that have been isolated from diverse habitats across six continents (Supplementary Fig. 2)[11,12]. We found significant variation in the ascr#5 response with high broad-sense heritability (0.83), indicating that most of the observed phenotypic variance can be explained by genetic differences across these strains (Fig. 2a). The two strains that represent the phenotypic extremes of the ascr#5 response are EG4349 and JU2576, where EG4349 did not enter dauer and was completely insensitive to ascr#5 treatment, and a large fraction of the JU2576 individuals entered the dauer stage in the same condition. Overall, we observed a continuous distribution of dauer-pheromone responses among these wild strains (mean = 0.41, standard deviation = 0.20), indicating that natural variation in this trait is likely not explained by a single gene.

To characterize the quantitative trait loci (QTL) associated with variation in the ascr#5 response, we performed genome-wide association (GWA) mappings and identified four QTL (Fig. 2b, c). The QTL that explained the most variation in pheromone-induced dauer induction (15.9%) is on the right arm of the X chromosome. Strains that have the non-reference (ALT) allele at the peak marker (X:14,145,335) of this QTL were less responsive to ascr#5 treatment than strains that have the reference (REF) allele (REF mean: 0.46; ALT mean: 0.30, $\log_{10}p$ = -5.851505). The remaining QTL on chromosomes II, III, and IV, explain 8.4%, 15.1%, and 5.4% of the variation in the ascr#5 response, respectively. We also found no obvious linkage disequilibrium (LD) among these QTL (Supplementary Fig. 3), suggesting that multiple genomic loci underlie natural variation in the ascr#5 response.

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

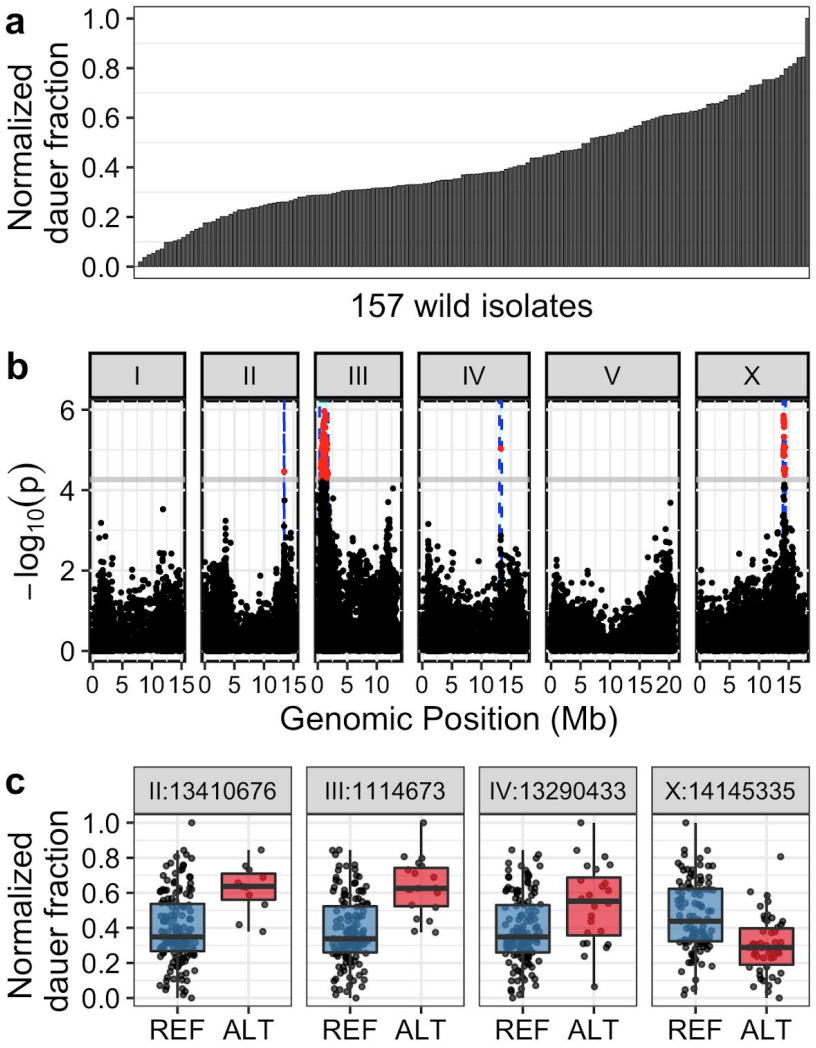

Fig. 2: Genome-wide association (GWA) mapping reveals four major loci underlying natural variation in dauer-pheromone response

(a) A bar plot for the natural variation of ascr#5-induced dauer formation at 25ºC across 157 *C. elegans* wild isolates (one-way analysis of variance (ANOVA), $\log_{10}p$ = -49.6598) is shown. Each bar represents the phenotypic response of a single wild isolate to 800 nM ascr#5. (b) A manhattan plot for single-marker based GWA mapping of the ascr#5-induced dauer formation trait from (a) is shown. Each dot represents a single-nucleotide variant (SNV) that is present in at least 5% of the 249 wild strains. The genomic position in Mb, separated by chromosome, is plotted on the x-axis, and the statistical significance of the correlation between genotype and phenotype is plotted on the y-axis. SNVs are colored red if they pass the genome-wide eigen-decomposition significance threshold, which is denoted by the gray horizontal line. The region of interest for each QTL is represented by vertical blue dotted lines. (c) Tukey box plots of phenotypes split by peak marker position of the four QTL (chrII:13410676, chrIII:1114673, chrIV:13290433, chrX:14145335) are shown. Each dot corresponds to the phenotype of an individual strain, which is plotted on the y-axis as the normalized dauer fraction phenotype. Strains are grouped by their genotype at each peak QTL position, where REF (blue) corresponds to the reference allele from the laboratory N2 strain and ALT (red) corresponds to the alternative allele.

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

A putative loss-of-function allele in an ascr#5 receptor gene is associated with reduced dauer formation

We focused our efforts on the largest effect QTL, which we named *dauf-1* (dauer-formation QTL #1). The 469-kb surrounding the *dauf-1* peak marker contains 82 protein-coding genes, including the duplicated genes *srg-36* and *srg-37*, which encode ascr#5 receptors[13]. Both genes are expressed in the same pair of chemosensory neurons (ASI), which play an essential role in the dauer-pheromone response[14,15]. Notably, previous studies reported that these two genes are repeatedly eliminated during long-term propagation of *C. elegans* in high-density liquid cultures[13]. Large deletions that remove both *srg-36* and *srg-37* arose in two independent laboratory-domesticated lineages, causing the loss of sensitivity to ascr#5.

To evaluate whether similar mutations in these two genes underlie the *dauf-1* QTL, we investigated the genome sequences of 249 wild strains available through the *C. elegans* Natural Diversity Resource (CeNDR)[16,17]. Although we could not find a large deletion that removes both *srg-36* and *srg-37*, we found one strain with a 411-bp deletion in *srg-36* and many other strains with an identical 94-bp deletion in *srg-37* (Fig. 3a, Supplementary Fig. 4). We named these deletions *srg-36(ean178)* and *srg-37(ean179)*. To test whether these deletions can explain the *dauf-1* QTL effect, we analyzed the association between the ascr#5 response and the two deletions. First, we found that *srg-36(ean178)*, which is a deletion found only in the PB303 strain and removes the fourth and fifth exons, is associated with an insensitivity to a high dose of ascr#5 (2 μM) (Supplementary Fig. 5). Because this deletion allele was not found in any other wild strains, *srg-36(ean178)* cannot explain the population-wide differences in dauer formation. By contrast, we found that all wild strains with the *srg-37*(*ean179)* deletion belong to *dauf-1(ALT)* group and had reduced ascr#5 sensitivity (Fig. 3b), suggesting that this deletion allele might cause a reduction in the acr#5 response.

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

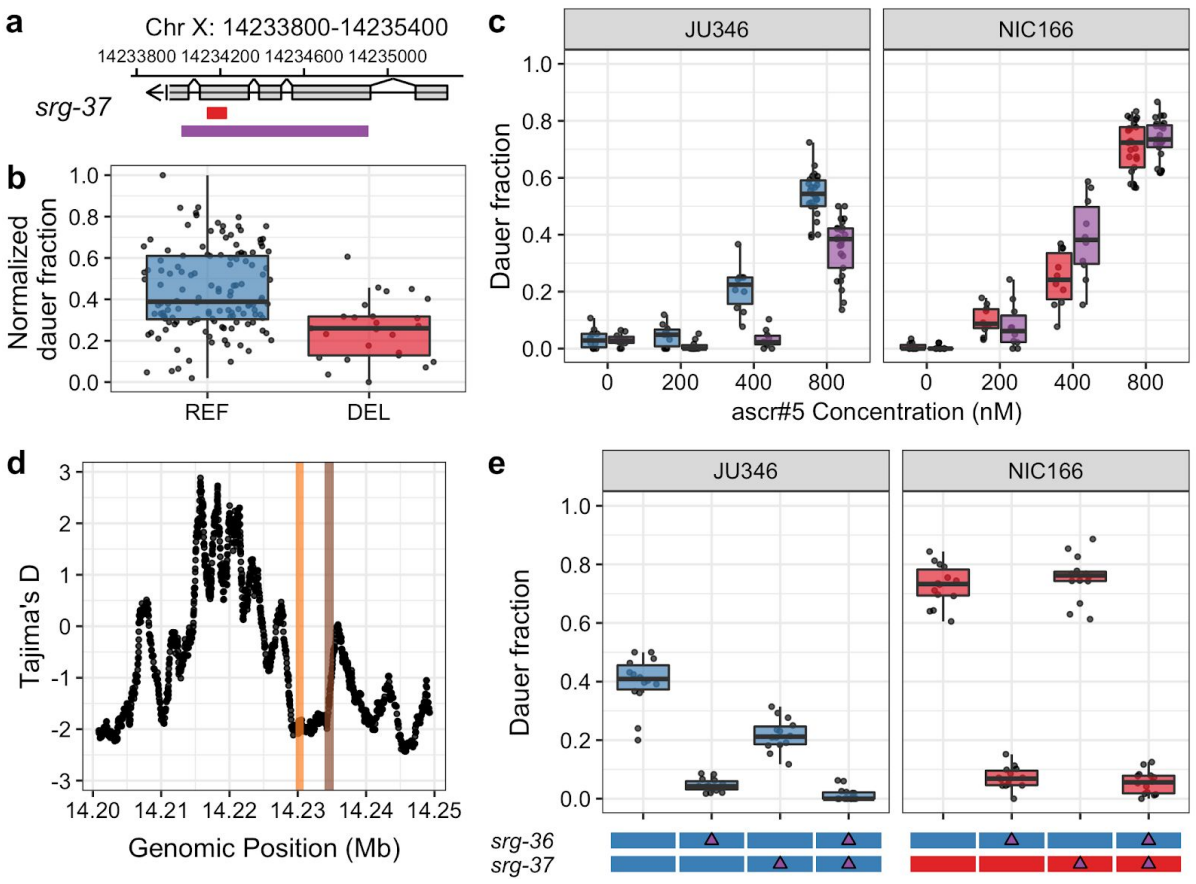

Fig. 3: Natural variant in the ascr#5 receptor gene, *srg-37*, underlies natural differences in dauer formation

(a) A schematic plot for the *srg-37* gene structure (grey), 94-bp natural deletion allele *ean179* (red), and CRISPR-Cas9 genome-editing target sequences for the putative loss-of-function deletion (purple) are shown. (b) Tukey box plots of dauer formation split by *srg-37* genotype are shown. Each dot corresponds to the phenotype of an individual strain, which is plotted on the y-axis by the normalized dauer fraction. Strains are grouped by their *srg-37* genotype, where REF (blue) corresponds to the wild-type reference allele from the laboratory N2 strain and DEL (red) corresponds to the natural 94-bp deletion allele (*ean179*). (c) Tukey box plots of the ascr#5 dose-response differences at 25ºC among two wild isolates and *srg-37(lf)* mutants in both backgrounds are shown with data points plotted behind. A dose response comparison is shown between (Left) JU346 *srg-37(+)* (blue) and JU346 *srg-37(lf)* (purple) and (Right) NIC166 *srg-37(ean179)* (red) and NIC166 *srg-37(lf)* (purple). The concentration of ascr#5 is shown on the x-axis, and the fraction of dauer formation is shown on the y-axis. (d) Tajima's D statistics across the *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). (e) Tukey box plots of *srg-36* and *srg-37* loss-of-function experiments under control (red, 0.4% ethanol) and ascr#5 pheromone conditions (blue, 2 µM of ascr#5) at 25ºC are shown with data points plotted behind. Genotypes of *srg-36* and *srg-37* are shown as colored bars on the x-axis, where red bar corresponds to the genotype with the natural deletion allele, *srg-37(ean179),* and the blue bar corresponds to the genotype without the deletion. Purple triangles represent the CRISPR-Cas9-mediated loss-of-function mutation. Fractions of dauer formation are shown on the y-axis.

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

The *srg-37*(*ean179)* deletion removes 31 amino acids surrounding the pocket structure of the G protein-coupled receptor and causes a frameshift mutation for the 46 C-terminal amino acids, together removing 23% (77/324) of the predicted SRG-37 amino acid sequence. Thus, this deletion likely impairs SRG-37 function, which could cause lower ascr#5 sensitivity. We hypothesized that, if *srg-37(ean179)* causes loss of gene function, removal of additional *srg-37* coding sequence would not further reduce the ascr#5 sensitivity of *srg-37(ean179)* wild strains. Using CRISPR-Cas9 genome-editing[18,19], we removed most of the coding sequences of *srg-37* from wild strains with both wild-type (reference-like) *srg-37* and the natural *srg-37* deletion (Fig. 3a). Indeed, we observed that a large deletion in *srg-37* did not change the ascr#5 sensitivities of two wild isolates with the natural deletion, but reduced the ascr#5 sensitivities of five wild isolates with reference-like *srg-37* (Fig. 3c, Supplementary Fig. 6), indicating that the natural deletion is likely a loss-of-function allele. Taken together, these results show that deletion of an ascr#5 receptor gene underlies natural variation in the dauer-pheromone response across the *C. elegans* population.

Selection has shaped the genetic variation of the two duplicated *C. elegans* ascr#5-receptor genes

We performed population genetic analysis across the *srg-36* and *srg-37* region by analyzing the genome sequences of 249 wild strains. Natural selection and demographic change can shift the allele frequency spectrum from neutrality, as measured by Tajima's D[20]. Purifying selection, a selective sweep, or a recent population expansion can cause accumulation of rare alleles at a given locus, indicated by a negative Tajima's D value. We found that the Tajima's D values were lowest across the promoter and coding regions of *srg-36* and increase back to background neutrality rates in the promoter region of *srg-37* (Fig. 3d). To specifically dissect selection features of *srg-36* and *srg-37*, we further analyzed the ratio between the non-synonymous substitution rate (Ka) and the synonymous

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

substitution rate (Ks) of the two genes by comparing nine homologs across four *Caenorhabditis* species[21]. We found that the non-synonymous substitution rate is much lower than the synonymous substitution rate (Ka/Ks < 1) for both genes, and the Ka/Ks statistic of *srg-36* (0.3145) is lower than *srg-37* (0.4223). Differences in deletion allele frequencies between *srg-36* and *srg-37* also correlate with stronger purifying selection at *srg-36*. The 411-bp deletion allele, *srg-36(ean178)*, is only found in a single wild isolate (PB303), whereas 18.4% (46/249) of wild strains carry the 94-bp deletion allele, *srg-37(ean179)*. These results together suggest that the two pheromone receptor genes are both under purifying selection, but the selection pressure is stronger for *srg-36*.

Although *srg-36* and *srg-37* are duplicated genes that are specific to the same ligand and expressed in the same cells, differences in non-coding and coding sequences between the two genes can cause differences in gene expression levels and receptor activities. Notably, previous studies report that transgene expression of *srg-36* showed a stronger effect than *srg-37* on the ascr#5 response[13]. To test whether *srg-36,* which is likely under stronger purifying selection than *srg-37*, plays a larger role in the ascr#5 response, we performed loss-of-function experiments. We removed the entire *srg-36* coding region in two wild strains – JU346 with wild-type (reference-like) *srg-37* and NIC166 with the natural *srg-37* deletion (Supplementary Fig. 4). First, we found that *srg-36(lf)* reduced ascr#5 sensitivity of both strains, indicating that *srg-36* is functional in both genetic backgrounds (Fig. 3e). Second, we observed that loss of *srg-36* reduced ascr#5 sensitivity more than loss of *srg-37*, supporting the conclusion that *srg-36* plays a larger role than *srg-37* in the ascr#5 response.

The higher activity of *srg-36* could be explained by differences in gene expression levels. We investigated the relative levels of *srg-36* and *srg-37* at the L1 stage, when these genes play critical roles in the dauer-pheromone response, and found that the expression levels of both genes are not significantly different[22]. It is more likely that differences in

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

protein-coding sequences cause the functional differences in the ascr#5 response. Although SRG-36 and SRG-37 show similarities in size and transmembrane structures (Supplementary Fig. 7)[23], only 46.4% of the amino acid residues are conserved between both receptors. The molecular differences between the two ascr#5 receptors could cause quantitative differences in ascr#5-receptor activities. Taken together, we hypothesized that *srg-36* is the primary ascr#5-receptor gene and is maintained across the *C. elegans* species through purifying selection. By contrast, the redundancy of these two genes facilitates *srg-37* variation and a loss-of-function allele can arise and spread across the population.

The *srg-37* deletion has spread globally and outcrossed with diverse genotypes

We investigated the locations where wild strains with the natural *srg-37* deletion were isolated and found 46 wild strains with this allele were isolated from all six continents (Fig. 4a, b). Given the low probability of acquiring the same 94-bp deletion, we hypothesized that this allele did not independently arise across multiple global locations but originated from a single ancestral population and spread throughout the world. To test this hypothesis, we analyzed the haplotype composition of *C. elegans* wild isolates across the X chromosome. We reproduced previous studies that showed a recent global selective sweep on the X chromosome[24], and the shared haplotype on the X chromosome we call the swept haplotype here (Supplementary Fig. 8, 9). Notably, we found that all 46 strains with the *srg-37* deletion exclusively share the swept haplotype at the *srg-37* locus (Fig. 4c). This result not only demonstrates that this allele arose at a single location, but also implies that it has spread throughout the world along with the recent selective sweep. Because the *srg-37* locus is far from the most swept part of the X chromosome, many strains must have outcrossed, suggesting that *srg-37* is unlikely the driver of the X chromosome sweep. Specifically, we found that 34.1% (85/249) of wild strains have an X chromosome that is swept more than

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

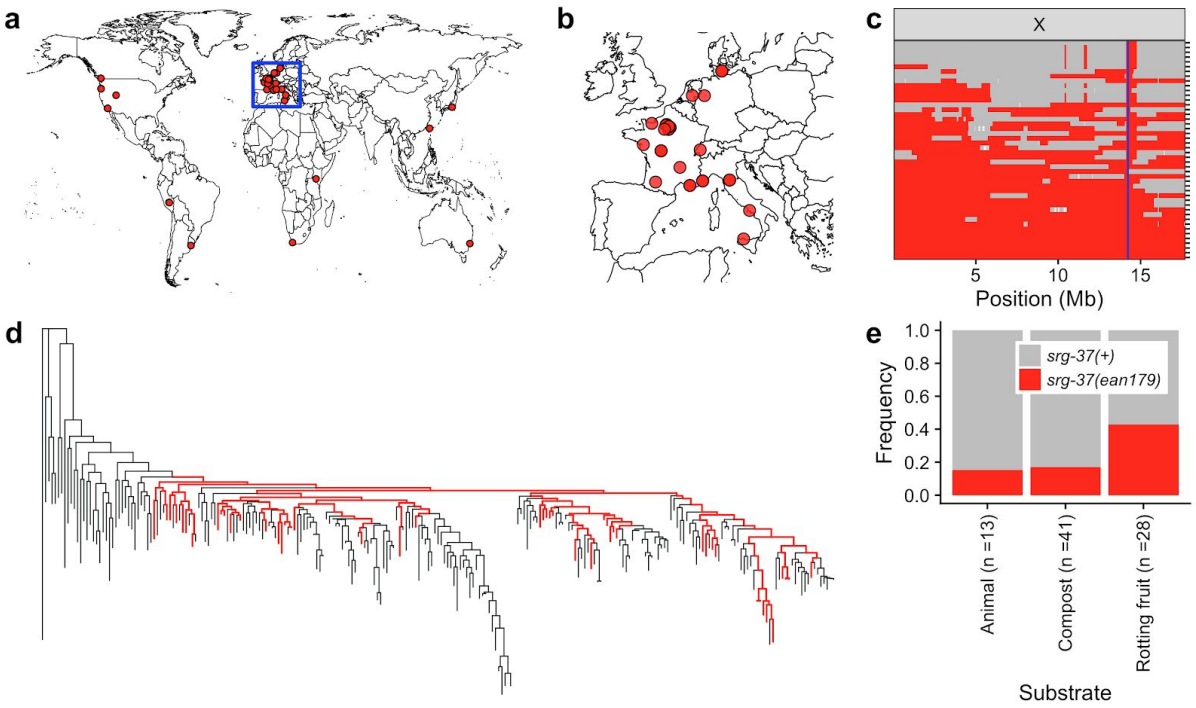

Fig. 4: Worldwide and niche-associated gene flow shape the ascaroside (ascr#5) pheromone receptor locus

(a) The global distribution of wild strains that contain the *srg-37(ean179)* deletion allele (red circle) is shown. (b) The geographic distribution of wild strains that are sampled from Europe (inset). Wild strains that contain the *srg-37* deletion (red circle) are shown. (c) Sharing of the swept haplotype on the X chromosome among 46 wild isolates with *srg-37(ean179)* is shown. Each row is one of the 46 strains, ordered roughly by the extent of swept-haplotype sharing (red). Other haplotypes are colored grey. Genomic position of X chromosome is shown on the x-axis. The blue line shows the position of the *srg-37* locus. (d) The genome-wide tree of 249 *C. elegans* wild isolates with those strains that have the *srg-37* deletion shown as red. (e) Stacked bar plots of *srg-37(+)* (grey) and *srg-37(ean179)* (red) allele frequencies among three subpopulations that were sampled from different substrates across hybrid zone in Europe (see Materials and methods). Substrate types and sample sizes are shown on the x-axis, and allele frequencies of each allele are shown on the y-axis. The *srg-37(ean179)* deletion allele is significantly enriched in the rotting fruit subpopulation (p = 0.0026).

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

50% of its length but have diverse non-swept haplotypes at the *srg-37* locus (Supplementary Fig. 10). These results suggest that wild strains have undergone outcrossing of the *srg-37* locus, which in turn purged the *srg-37* deletion from these swept subpopulations (Fig. 4d). Taken together, the results suggest that the *srg-37* deletion spread globally with the selective sweeps but was lost because of more recent outcrossing.

The *srg-37* deletion is enriched in the rotting fruit niche

Dauer development is a trade-off between long-term survival and short-term reproductive success. Because reduction of the dauer-pheromone response can promote reproductive growth, we investigated whether wild strains with the *srg-37* deletion were sampled more often from substrates with proliferating populations. These populations are often found in nutritious habitats, such as rotting vegetation[25]. By contrast, *C. elegans* were sampled predominantly in the dauer stage from animal and compost substrates[6,26]. We analyzed the allele frequencies of the *srg-37* deletion among three subpopulations sampled from animals, compost, and rotting fruits across geographic locations where both *srg-37* alleles were isolated (Supplementary Fig. 11, see Materials and methods). We found that wild strains with the *srg-37* deletion were 67% enriched in rotting fruits (Fig. 4e, Supplementary Table 1). Thus, this allele is not only associated with lower dauer-pheromone responses but also with natural substrates that are known to support reproductive growth.

Discussion

Dauer pheromones are chemical signals that are perceived by sensory neurons through the combined actions of chemoreceptors and cGMP-mediated signaling[5,27]. In the absence of dauer-pheromone signaling, the insulin/IGF-1 and TGF-β signaling pathways promote reproductive growth through the production of steroid hormones (dafachronic acid)[28]. Genetic variation in the genes that mediate pheromone perception or downstream

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

signaling likely alter an individual's dauer-pheromone response. However, because the signaling pathways that act downstream of pheromone perception are involved in various biological processes[29,30], mutations in these pathways might cause deleterious pleiotropic effects. Previous studies have shown that the ascr#5-receptor SRG-37 was lost in two independent laboratory lineages of *C. elegans*[13], suggesting that selection more readily acts at the pheromone perception step of this developmental pathway. In this study, we provide further support for this hypothesis by showing that 18% of wild *C. elegans* strains harbor a putative loss-of-function deletion in the ascr#5-receptor SRG-37, and that these individuals are more likely to be found in nutrient-rich habitats. Thus, modification of pheromone-receptor activity might be favored in both laboratory and natural conditions to fine-tune dauer-pheromone responses with few pleiotropic effects[13,31]. However, we identified additional dauer-pheromone response QTL, suggesting that multiple loci are involved in ascr#5 responses. Interestingly, SRG-36 and SRG-37 are the only two known ascr#5 receptors involved in dauer-pheromone signaling. Therefore, the other three ascr#5-response QTL indicate that additional natural genetic variants could affect uncharacterized ascr#5 receptors, novel or known factors that regulate receptor activity, or downstream signaling components.

Insights into the redundant functions of *srg-36* and *srg-37* were first gained from the observation that both genes were deleted from two independent laboratory-domesticated *C. elegans* lineages[13]. We did not find a single wild strain in the *C. elegans* population that carries a deletion of both *srg-36* and *srg-37*. Investigations of within (Tajima's *D)* and among (Ka/Ks) species neutrality statistics suggest that selection acts on these two genes differently. Our results imply that the *srg-36* and *srg-37* genes might not be functionally equivalent in the wild population. The loss-of-function experiments suggest that *srg-36* plays a larger role in the ascr#5 response than *srg-37.* Given the important role of the dauer stage in the long-term survival and the phoretic dispersal of the species, purifying selection might

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

act to conserve the primary ascr#5 receptor (SRG-36) in the *C. elegans* population to maintain the dauer-pheromone response.

The gene-dosage hypothesis proposes that duplicated genes can each become fixed in a population if a dosage increase causes a selective advantage[32]. Recent studies in the natural budding yeast population have reported that copy number variants (CNVs) can explain larger proportions of trait variance compared to single-nucleotide variants (SNVs)[33]. It was also shown that more than 30% of human olfactory-receptor genes exhibit copy number variation[34], so signal perception might be a fruitful substrate for trait evolution. We show that dosage differences of the ascr#5 receptor genes explain the largest dauer-pheromone response QTL, suggesting a pervasive effect of CNVs across species. Our observation that the *srg-37* deletion is enriched in rotting fruits provides evidence that selection can act on CNVs in a niche-specific manner. Although our studies focus on the ascr#5 receptors, approximately 1,300 chemoreceptor genes and 400 pseudogenes are annotated in the reference *C. elegans* genome[35,36]. By contrast, parasitic nematodes tend to have fewer chemoreceptors[37], and these species inhabit less heterogeneous niches. Therefore, we hypothesize that the expanded diversity of chemoreceptors and the prevalence of chemoreceptor pseudogenes in free-living *Caenorhabditis* species is caused by broad context-dependent selection in diverse niches. Furthermore, the diversity of the chemoreceptor repertoires is captured across wild *C. elegans* strains that are sampled from various niches[38,39]. Taken together, these observations indicate that niche adaptation could facilitate the dynamic duplication, deletion, and functional diversification of chemoreceptor gene evolution.

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Materials and methods

C. elegans strains

Animals were cultured at 20ºC on modified nematode growth medium (NGMA) seeded with the *E. coli* strain OP50[40]. Prior to each assay, strains were passaged for at least four generations without entering starvation or encountering dauer-inducing conditions. For the genome-wide association (GWA) studies, 157 wild isolates from CeNDR (version 20170531) were used[16,17]. All strain information can be found in Supplementary Table 2.

High-throughput dauer assay

Strains were propagated for four generations on agar plates, followed by bleach synchronization. Approximately 50 embryos were titered and placed into each well of a 96-well microtiter plate filled with 50 µL of K medium[41] with modified salt concentrations (10.2 mM NaCl, 32 mM KCl, 3 mM $CaCl_2$, 3 mM $MgSO_4$), 50 µM kanamycin, 5 mg/mL HB101 bacterial lysate (Pennsylvania State University Shared Fermentation Facility, State College, PA), and synthetic ascaroside[42] dissolved in 0.4% ethanol or 0.4% ethanol alone. Animals were cultured for 52 hours at 25ºC until they reached the young adult stage or arrested at the dauer stage. Animals were exposed to 0.5 µm fluorescent microspheres (Polysciences, cat. # 19507-5) at a final concentration of 7.28 x 10^8 particles/mL and 5 µL of 1 mg/mL HB101 bacterial lysate to promote feeding for 20 minutes. After this exposure, 200 µL of 50 mM sodium azide was added to each well to kill the animals, stop feeding, and straighten the animals. Using the COPAS BIOSORT large particle flow cytometer (Union Biometrica, Holliston MA), optical parameters of animals, including fluorescence intensity, time-of-flight (TOF, animal length), and extinction (optical density) were measured. Measured parameters were used to build a model that can differentiate dauer and adult stages of the population in each well through the R package EMCluster[43]. One cluster with lower fluorescence and

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

smaller body size was assigned to the dauer population and the other to the non-dauer population. The dauer fraction was calculated per well as a fraction of dauer animals among total animals, which is shown as a single data point in each plot. From the control experiments, both the false positive ratio (false dauer detection in a wild-type sample without pheromone treatment) and the false negative ratio (false non-dauer detection in Daf-c mutant sample) were 5%, indicating 95% accuracy of the assay (Fig. 1b,c).

Heritability calculations

For dose response experiments, broad-sense heritability (H^2) estimates were calculated using the *lmer* function in the lme4 package with the following linear mixed model (phenotype ~1 + (1|strain))[44]. H^2 was then calculated as the fraction of the total variance that can be explained by the random component (strain) of the mixed model.

Genome-wide association mapping

A genome-wide association (GWA) mapping was performed using phenotype data from 157 wild *C. elegans* strains. Genotype data were acquired from the latest VCF release (Release 20180527) from CeNDR that was imputed as described previously[16]. We used BCFtools[45] to filter variants that had any missing genotype calls and variants that were below 5% minor allele frequency. We used PLINK v1.9[46,47] to LD-prune the genotypes at a threshold of $r^2 <$ 0.8, using *--indep-pairwise 50 10 0.8.* The pruned genotype set comprised 72,568 markers that were used to generate the realized additive kinship matrix using the *A.mat* function in the *rrBLUP* R package[48]. These markers were also used for genome-wide mapping. However, because these markers still have substantial LD within this genotype set, we performed eigen decomposition of the correlation matrix of the genotype matrix using *eigs_sym* function in Rspectra package[49]. The correlation matrix was generated using the *cor* function in the correlateR R package[50]. We set any eigenvalue greater than one from this

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

analysis to one and summed all of the resulting eigenvalues[51]. This number was 915.621, which corresponds to the number of independent tests within the genotype matrix. We used the *GWAS* function in the rrBLUP package to perform genome-wide mapping with the following command: *rrBLUP::GWAS(pheno = dauer, geno = Pruned_Markers, K = KINSHIP, min.MAF = 0.05, n.core = 1, P3D = FALSE, plot = FALSE)*. Regions of interest are defined as +/- 100 SNVs from the rightmost and leftmost markers above the eigen-decomposition significance threshold. If regions of interest for separate QTL are within 1000 SNVs, they become grouped as a single region of interest.

Identification of natural deletion variants of *srg-36* and *srg-37*

Whole-genome sequence data were aligned to WS245 using bwa (version 0.7.8-r455). Optical/PCR duplicates were marked with PICARD (version 1.111)[16,52–54]. Alignments with greater than 100X coverage were subsampled to 100X using sambamba[55]. We called large deletions using the Manta structural variant caller (v1.4.0) using the default caller and filter settings[56].

Generation of *srg-36* and *srg-37* deletion strains

srg-36 and *srg-37* loss-of-function mutant strains were generated by CRISPR-Cas9-mediated genome editing, using a co-CRISPR approach and Cas9 ribonucleoprotein (RNP) delivery[18,19]. crRNAs synthesized by IDT (Skokie, IL) targeting *srg-36* (exon 1 and the 3' UTR) and *srg-37* (exon 2 and exon 5) were used to generate deletions. The injection mixture (10 µL) was prepared with 0.88 µL of 200 µM tracrRNA (IDT, Product #1072532), 0.88 µL of 100 µM crRNA1 (5' targeting) and crRNA2 (3' targeting), and 0.12 µL of 100 µM *dpy-10* crRNA (IDT) were mixed and incubated at 95ºC for five minutes. After cooling to room temperature, 2.87 µL of 60 µM Cas9 protein (IDT Product #1074181) was added and incubated at room temperature for five minutes. Finally, 0.5 µL of 10 µM

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

dpy-10 ssODN (IDT) repair template and 3.99 µL of nuclease free water were added. RNP injection mixtures were microinjected into the germline of young adult hermaphrodites (P0), and injected animals were singled to fresh 6 cm NGM plates 18 hours after injection. Two days later, F1 progeny were screened, and animals expressing a Rol phenotype were transferred to new plates and allowed to generate progeny (F2). Afterwards, F1 animals were genotyped by PCR. Deletion of *srg-36* was detected with primers oECA1460-1463 and deletion of *srg-37* was detected with primers oECA1429, oECA1430, and oECA1435. Non-Rol progeny (F2) of F1 animals positive for the desired deletion were propagated on separate plates to generate homozygous progeny. F2 animals were genotyped afterwards with same primer sets, and PCR products were Sanger sequenced for verification. All crRNA and oligonucleotide sequences are listed in the Supplementary Table 3.

Population genetics

Sliding window analysis of population genetic statistics was performed using the PopGenome package in R[57]. All sliding window analyses were performed using the imputed SNV VCF available on the CeNDR website with the most diverged isotype XZ1516 set as the outgroup[16,58,59]. Estimates of Ka/Ks were performed using an online service (http://services.cbu.uib.no/tools/kaks). Linkage disequilibrium (LD) of QTL markers, which can be measured as the square of the correlation coefficient (r^2), was calculated using the genetics package in R[60]. The formula for the correlation coefficient is $r = -D / sqrt\ (p(A) * p(a) * p(B) * p(b))$. Haplotype composition of each wild isolate was inferred by applying IBDseq[61] with variants called by BCFtools[62] and the following filters: Depth (DP) > 10; Mapping Quality (MQ) > 40; Variant quality (QUAL) > 10; (Alternate-allelic Depth (AD) / Total Depth (DP)) ratio > 0.5; < 10% missing genotypes; < 10% heterozygosity.

Substrate specificity analysis in the co-sampling zone

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

The co-sampling zone was defined as a location where both *srg-37(+)* and *srg-37(ean179)* were isolated (Supplementary Fig. 11). Collection information available on the CeNDR website were used to analyze correlation between isolated substrate and the *srg-37* genotype of each isolate. Isolation of wild strains that share same genome-wide genotypes (isotype) were counted as independent isolations if they were sampled from different locations or from different substrate types. We found that 95 isotypes were isolated in co-sampling zone from at least 119 independent isolations. Three substrates (animals, compost, rotting fruit) with more than ten independent isolated strains were selected for substrate enrichment test. In total, 82 wild strains (66 isotypes) were grouped into three subpopulations by the substrate where they were isolated, and allele frequencies of each subpopulation were calculated. Significant enrichment of *srg-37(ean179)* in each subpopulation was determined by hypergeometric tests using stats R package[63].

Acknowledgements

This work was supported by an NSF CAREER Award to E.C.A. S.Z was supported by the Cell and Molecular Basis of Disease training grant (T32GM008061) and The Bernard and Martha Rappaport Fellowship. D.E.C received the National Science Foundation Graduate Research Fellowship (DGE-1324585). Members of the Andersen Lab helped manuscript editing and Ying K. Zhang for assistance with the synthesis of ascarosides. Some strains were provided by the CGC, which is funded by the NIH Office of Research Infrastructure Programs (P40 OD010440). We want to thank WormBase for providing genome data of *Caenorhabditis* species.

Author contributions

D.L. and E.C.A. conceived and designed the study. D.L. performed the high-throughput assay, CRISPR-Cas9 genome-editing, population genomic analyses, and niche enrichment

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

tests. S.Z. performed the GWA mapping, identified genetic variants in the *dauf-1* locus, generated the genome-wide tree of 249 wild *C. elegans* strains, and edited the manuscript. D.E.C. analyzed the haplotype composition of 249 wild strains. L.F., J-C.H., M.G.S., J.A.G.R., J.W., J.E.K., C.B., and M-A.F. contributed wild isolates to the 249 wild *C. elegans* strain collection. F.C.S provided the dauer pheromone. D.L. and E.C.A. analyzed the data and wrote the manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

References

1. Cassada, R. C. & Russell, R. L. The dauerlarva, a post-embryonic developmental variant of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **46**, 326–342 (1975).
2. Golden, J. W. & Riddle, D. L. A pheromone influences larval development in the nematode *Caenorhabditis elegans*. *Science* **218**, 578–580 (1982).
3. Golden, J. W. & Riddle, D. L. The *Caenorhabditis elegans* dauer larva: developmental effects of pheromone, food, and temperature. *Dev. Biol.* **102**, 368–378 (1984).
4. Schaedel, O. N., Gerisch, B., Antebi, A. & Sternberg, P. W. Hormonal signal amplification mediates environmental conditions during development and controls an irreversible commitment to adulthood. *PLoS Biol.* **10**, e1001306 (2012).
5. Ludewig, A. H. & Schroeder, F. C. Ascaroside signaling in *C. elegans*. *WormBook* 1–22 (2013).
6. Félix, M.-A. & Braendle, C. The natural history of *Caenorhabditis elegans*. *Curr. Biol.* **20**, R965–9 (2010).
7. Frézal, L. & Félix, M.-A. *C. elegans* outside the Petri dish. *Elife* **4**, (2015).
8. Lee, H. *et al.* Nictation, a dispersal behavior of the nematode *Caenorhabditis elegans*, is regulated by IL2 neurons. *Nat. Neurosci.* **15**, 107–112 (2011).
9. Lee, D. *et al.* The genetic basis of natural variation in a phoretic behavior. *Nat. Commun.* **8**, 273 (2017).
10. Nika, L., Gibson, T., Konkus, R. & Karp, X. Fluorescent Beads Are a Versatile Tool for Staging *Caenorhabditis elegans* in Different Life Histories. *G3* **6**, 1923–1933 (2016).
11. Evans, K. S. *et al.* Correlations of Genotype with Climate Parameters Suggest *Caenorhabditis elegans* Niche Adaptations. *G3* **7**, 289–298 (2017).
12. Schulenburg, H. & Félix, M.-A. The Natural Biotic Environment of *Caenorhabditis elegans*. *Genetics* **206**, 55–86 (2017).
13. McGrath, P. T. *et al.* Parallel evolution of domesticated Caenorhabditis species targets pheromone receptor genes. *Nature* **477**, 321–325 (2011).
14. Bargmann, C. I. & Horvitz, H. R. Control of larval development by chemosensory neurons in *Caenorhabditis elegans*. *Science* **251**, 1243–1246 (1991).
15. Schackwitz, W. S., Inoue, T. & Thomas, J. H. Chemosensory neurons function in parallel to mediate a pheromone response in *C. elegans*. *Neuron* **17**, 719–728 (1996).
16. Cook, D. E., Zdraljevic, S., Roberts, J. P. & Andersen, E. C. CeNDR, the *Caenorhabditis elegans* natural diversity resource. *Nucleic Acids Res.* **45**, D650–D657 (2017).
17. Hahnel, S. R. *et al.* Extreme allelic heterogeneity at a *Caenorhabditis elegans* beta-tubulin locus explains natural resistance to benzimidazoles. *PLoS Pathog.* **14**, e1007226 (2018).
18. Kim, H. *et al.* A co-CRISPR strategy for efficient genome editing in *Caenorhabditis elegans*. *Genetics* **197**, 1069–1080 (2014).
19. Paix, A., Folkmann, A., Rasoloson, D. & Seydoux, G. High Efficiency, Homology-Directed Genome Editing in *Caenorhabditis elegans* Using CRISPR-Cas9 Ribonucleoprotein Complexes. *Genetics* **201**, 47–54 (2015).
20. Biswas, S. & Akey, J. M. Genomic insights into positive selection. *Trends Genet.* **22**, 437–446 (2006).
21. Li, W. H., Wu, C. I. & Luo, C. C. A new method for estimating synonymous and

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.* **2**, 150–174 (1985).

22. Gerstein, M. B. *et al.* Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330**, 1775–1787 (2010).
23. Chang, J.-M., Di Tommaso, P., Taly, J.-F. & Notredame, C. Accurate multiple sequence alignment of transmembrane proteins with PSI-Coffee. *BMC Bioinformatics* **13 Suppl 4**, S1 (2012).
24. Andersen, E. C. *et al.* Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. *Nat. Genet.* **44**, 285–290 (2012).
25. Félix, M.-A. & Duveau, F. Population dynamics and habitat sharing of natural populations of *Caenorhabditis elegans* and C. briggsae. *BMC Biol.* **10**, 59 (2012).
26. Barrière, A. & Félix, M.-A. High local genetic diversity and low outcrossing rate in *Caenorhabditis elegans* natural populations. *Curr. Biol.* **15**, 1176–1184 (2005).
27. Hu, P. J. Dauer. *WormBook* 1–19 (2007).
28. Lee, S. S. & Schroeder, F. C. Steroids as central regulators of organismal development and lifespan. *PLoS Biol.* **10**, e1001307 (2012).
29. Gumienny, T. L. & Savage-Dunn, C. TGF-β signaling in *C. elegans*. *WormBook* 1–34 (2013).
30. Murphy, C. T. & Hu, P. J. Insulin/insulin-like growth factor signaling in *C. elegans*. *WormBook* 1–43 (2013).
31. Gompel, N. & Prud'homme, B. The causes of repeated genetic evolution. *Dev. Biol.* **332**, 36–47 (2009).
32. Kondrashov, F. A. & Koonin, E. V. A common framework for understanding the origin of genetic dominance and evolutionary fates of gene duplications. *Trends Genet.* **20**, 287–290 (2004).
33. Peter, J. *et al.* Genome evolution across 1,011 Saccharomyces cerevisiae isolates. *Nature* **556**, 339–344 (2018).
34. Nozawa, M., Kawahara, Y. & Nei, M. Genomic drift and copy number variation of sensory receptor genes in humans. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 20421–20426 (2007).
35. Thomas, J. H. & Robertson, H. M. The Caenorhabditis chemoreceptor gene families. *BMC Biol.* **6**, 42 (2008).
36. Vidal, B. *et al.* An atlas of *Caenorhabditis elegans* chemoreceptor expression. *PLoS Biol.* **16**, e2004218 (2018).
37. International Helminth Genomes Consortium. Comparative genomics of the major parasitic worms. *Nat. Genet.* **51**, 163–174 (2019).
38. Stewart, M. K., Clark, N. L., Merrihew, G., Galloway, E. M. & Thomas, J. H. High genetic diversity in the chemoreceptor superfamily of *Caenorhabditis elegans*. *Genetics* **169**, 1985–1996 (2005).
39. Greene, J. S., Dobosiewicz, M., Butcher, R. A., McGrath, P. T. & Bargmann, C. I. Regulatory changes in two chemoreceptor genes contribute to a *Caenorhabditis elegans* QTL for foraging behavior. *Elife* **5**, (2016).
40. Andersen, E. C., Bloom, J. S., Gerke, J. P. & Kruglyak, L. A variant in the neuropeptide receptor npr-1 is a major determinant of *Caenorhabditis elegans* growth and physiology. *PLoS Genet.* **10**, e1004156 (2014).
41. Boyd, W. A., Smith, M. V. & Freedman, J. H. *Caenorhabditis elegans* as a model in

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

developmental toxicology. *Methods Mol. Biol.* **889**, 15–24 (2012).
42. Zhang, Y. K., Sanchez-Ayala, M. A., Sternberg, P. W., Srinivasan, J. & Schroeder, F. C. Improved Synthesis for Modular Ascarosides Uncovers Biological Activity. *Org. Lett.* **19**, 2837–2840 (2017).
43. Chen, W.-C., Maitra, R. & Melnykov, V. A Quick Guide for the EMCluster Package. *R Vignette, URL http://cran. r-project. org/package= EMCluster* (2012).
44. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models using lme4. *arXiv [stat.CO]* (2014).
45. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
46. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
47. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
48. Endelman, J. B. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *Plant Genome* **4**, 250–255 (2011).
49. Qiu, Y. *RSpectra*. (Github).
50. Bilgrau, A. E. *correlateR*. (Github, 2018).
51. Li, J. & Ji, L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* **95**, 221–227 (2005).
52. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
53. Picard Tools - By Broad Institute. Available at: http://broadinstitute.github.io/picard/.
54. Cook, D. E. *et al.* The Genetic Basis of Natural Variation in *Caenorhabditis elegans* Telomere Length. *Genetics* **204**, 371–383 (2016).
55. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).
56. Chen, X. *et al.* Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).
57. Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E. & Lercher, M. J. PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.* **31**, 1929–1936 (2014).
58. Browning, B. L. & Browning, S. R. Genotype Imputation with Millions of Reference Samples. *Am. J. Hum. Genet.* **98**, 116–126 (2016).
59. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
60. Warnes G, Gorjanc G, Leisch F, Man M. *Genetics: Population genetics, R package version 1.3.6*. (2012).
61. Browning, B. L. & Browning, S. R. Detecting identity by descent and estimating genotype error rates in sequence data. *Am. J. Hum. Genet.* **93**, 840–851 (2013).
62. Danecek, P., Schiffels, S. & Durbin, R. Multiallelic calling model in bcftools (-m).
63. stats package | R Documentation. Available at: https://www.rdocumentation.org/packages/stats/versions/3.5.2.

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Supplementary Figures:

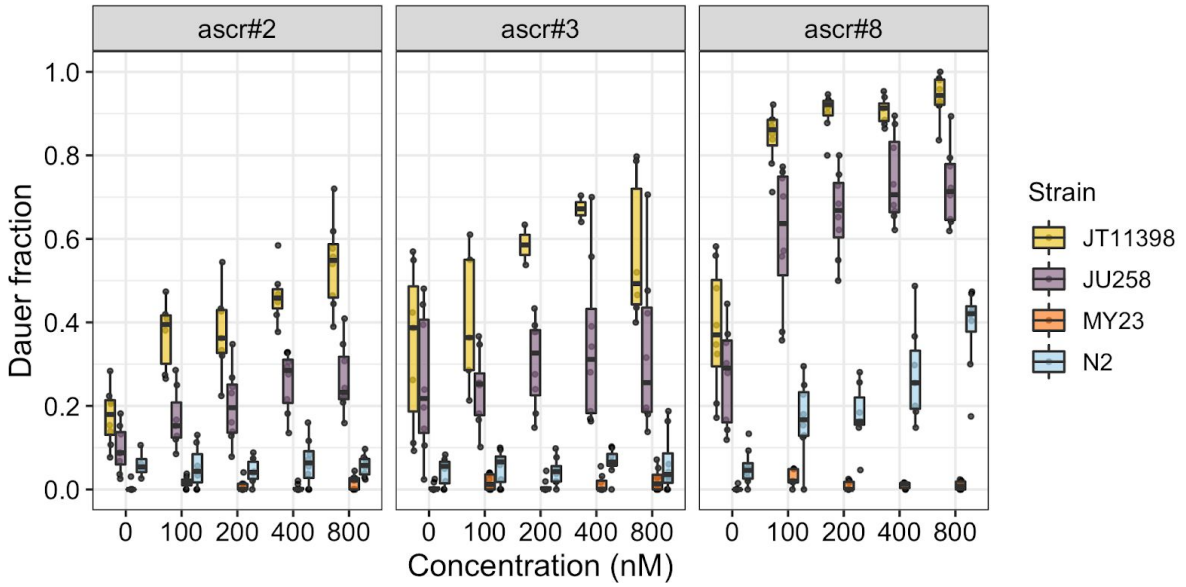

Fig. S1: Genetically divergent wild isolates show different dose responses to dauer pheromone (ascr#2, ascr#3, and ascr#8) treatments
Tukey box plots of the dose responses at 25ºC for ascr#2 (left), ascr#3 (center), and ascr#8 (right) for four divergent strains are shown with data points plotted behind. 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.

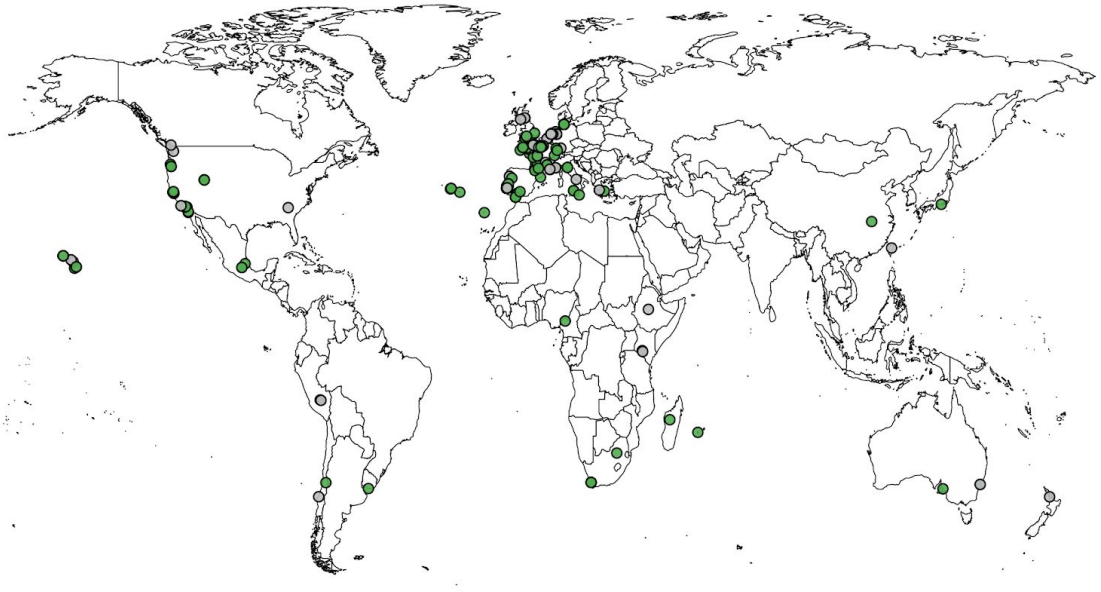

Fig. S2: Wild *C. elegans* strains have isolated across six continents
Among the 249 wild strains that are available through the CeNDR, the global distribution of the 239 wild strains with known geographic origins is shown. Among the 157 wild strains that were tested using the HTDA, 151 wild strains with known geographic origins are shown as green circles. Wild strains that are not tested using the HTDA are shown as grey circles.

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

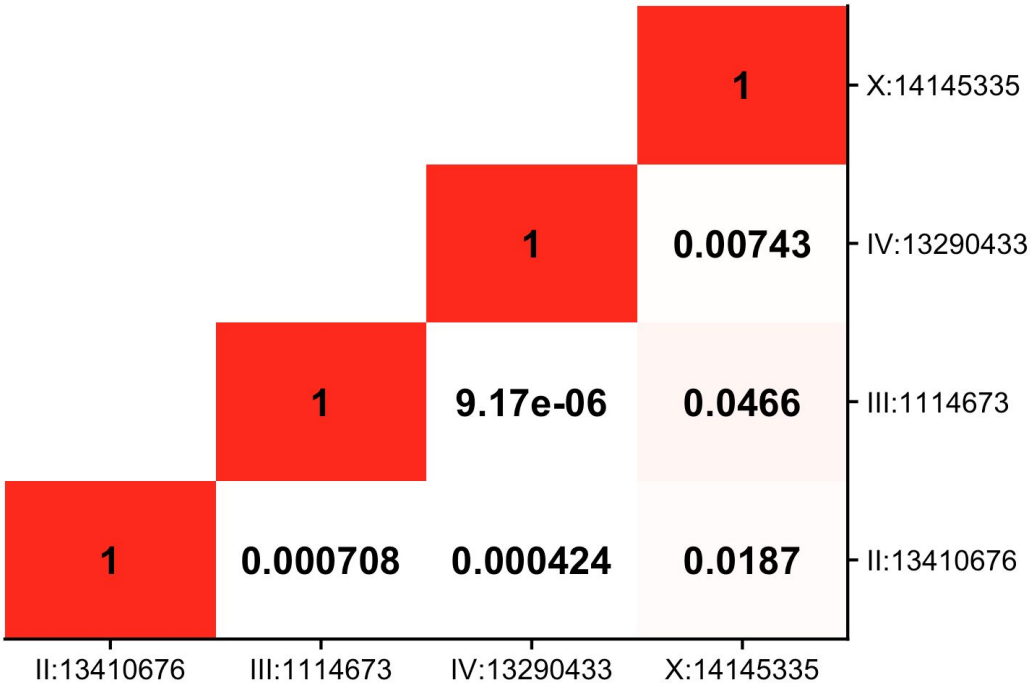

Fig. S3: Linkage disequilibrium is not observed among four QTL for ascr#5 response
A heatmap plot that shows linkage disequilibrium (LD) among peak QTL markers as measured by the square of the correlation coefficient (r^2).

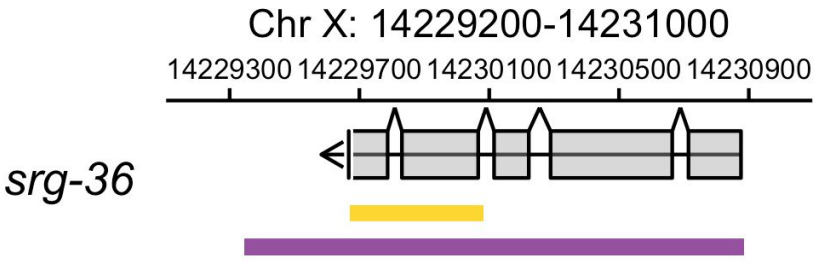

Fig. S4: A schematic plot for *srg-36* gene structure
A schematic plot for the *srg-36* gene structure (grey), the 411-bp natural deletion allele *ean178* (yellow), and the CRISPR-Cas9 loss-of-function deletion (purple) are shown.

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

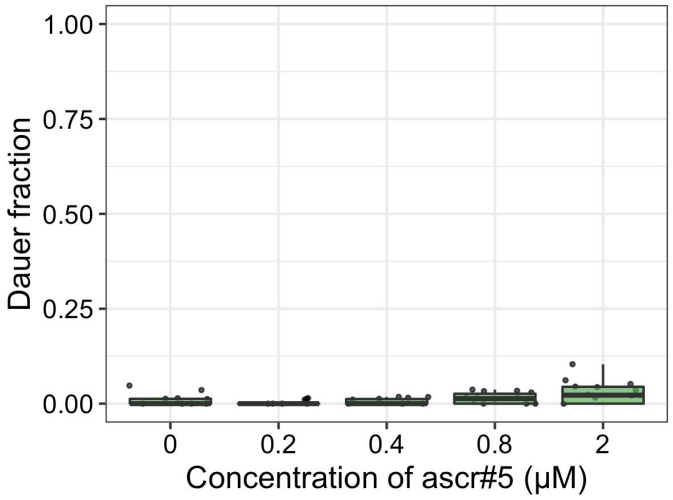

Fig. S5: PB303, a unique strain that has a deleted *srg-36*, is insensitive to ascr#5
Tukey box plots of the ascr#5 dose response at 25ºC for PB303 that carries the *srg-36(ean178)* deletion allele are shown with data points plotted behind. Concentrations of ascr#5 are shown on the x-axis, and the fraction of dauer formation is shown on the y-axis.

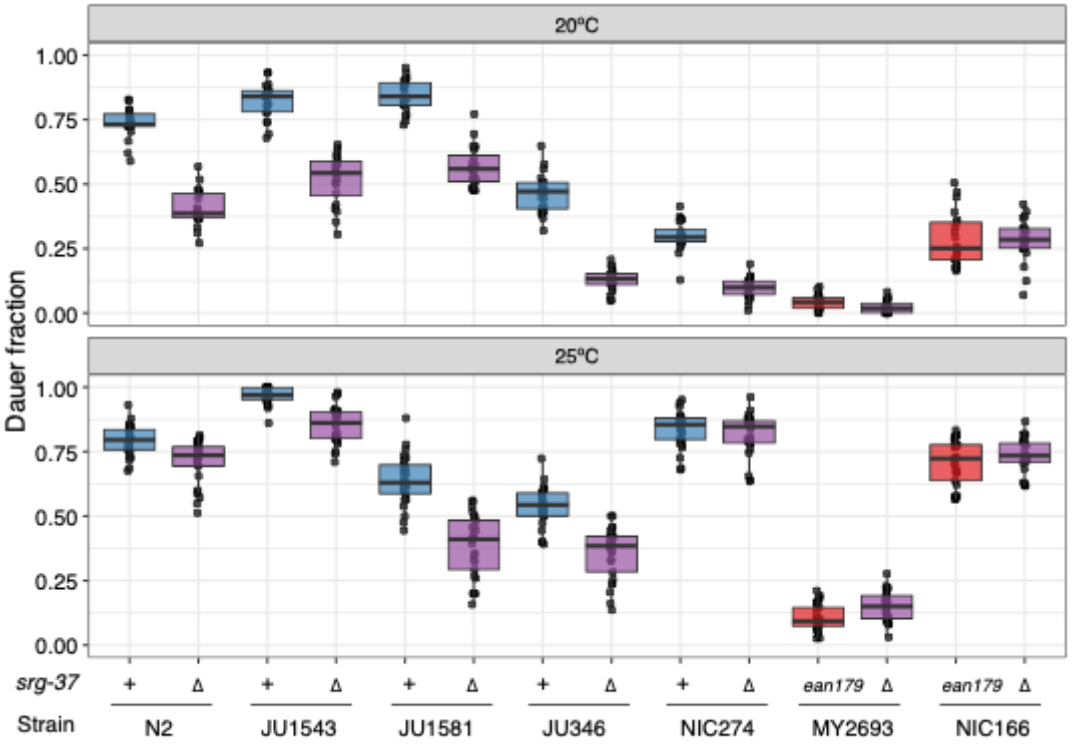

Fig. S6: The *srg-37(ean179)* deletion is a putative loss-of-function allele
Tukey box plots of *srg-37* loss-of-function experiments in dauer pheromone conditions (800 nM of ascr#5) at 20ºC (top) and 25ºC (bottom) are shown with data points plotted behind. Each box plot is colored by the genotypes of *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.

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

```
 IN  HEL  OUT

SRG-36   MTLASLLRFIITSVYGAALLVLYTYVVIIIIAHKKSESSFRSSFYKLFIVGFFMNMMTYFNSLISLRLPQSN
SRG-37   MPSSSPLRLIISTIYGIVCLLLYTLVTAVICA---SSSMIGSPFYRLFAIGYIMNILTYLNSFISIRLPQNT
         *. :* **:**:::** . *:*** *. :* *   *.* : *.**:** :*::**::**:**:**:****..

SRG-36   GINETLSNFFLTHNEHNMEVIFPVKVFHFLHYYFAYAQYIYNFFISLNRFSLITNPIKSELFWRKRFWWFVF
SRG-37   GINGNLSNFFLTHNEFNMNMGCPINLFHALHYYFAYTQYIYNFLISYNRFCAITSLLDIEKRWKRSLWIFVT
         *** .**********.**::  *:::** *******:******:** ***. **. :. *  *:: :* **

SRG-36   AIFAIPWAFAVPILVGRSYYVYHSHGDYFFLDTTIKRSLIYTILAPTLSVITAMNSFLNFMSVRKVIILKLS
SRG-37   IMFVYPMIPAGLIFFNRSYYTYVSSKDNFYISSTLGHKAIYGFLAPNLFIITIFNTVLNIKAYKKLLVLKKT
          :*. *   *  *:..****.* *  * *::.:*: :. ** :***.* :** :*:.**: : :*:::** :

SRG-36   GGRVPEKNLLQMSFVIFFIDIFLVILTLIKAIIIACGLSFKSESTDEILSWIVILIPFASDALTLTQPLLLL
SRG-37   LRSVPDTNLFYMSLAMFGIDSFLAVLSTFKAIIFLLELKNSSEFFERLVDWIDLLVPFASDALTLTQPLLLL
            **:.**: **:.:* ** **.:*: :****:   *. .**  :.::.** :*:****************

SRG-36   FFSKTVRRCCIWPFPCLKSLRSHRLIARNSVLVVRPFGAPTITG
SRG-37   YFSSTLRQKCIERLPFLKVFLNNRFFVRSRNMHQELTKY-----
         :**.*:*: **  :* ** : .:*::.*.  :  .
```

Fig. S7: 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[23]. 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).

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

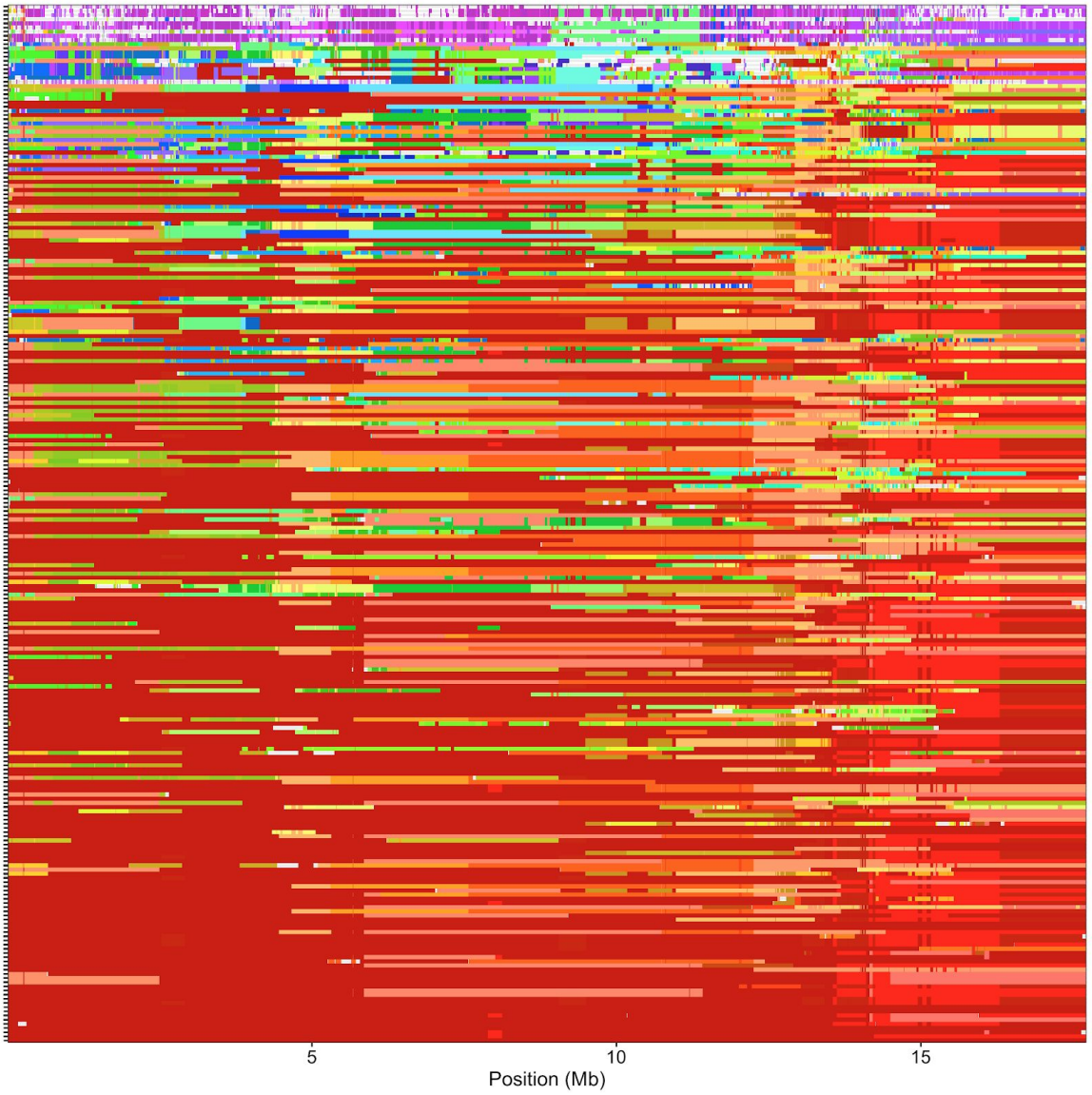

Fig. S8: X chromosome sharing of 249 wild strains reveals swept regions where haplotype diversity is reduced

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 wild strains (isotypes), 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.

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

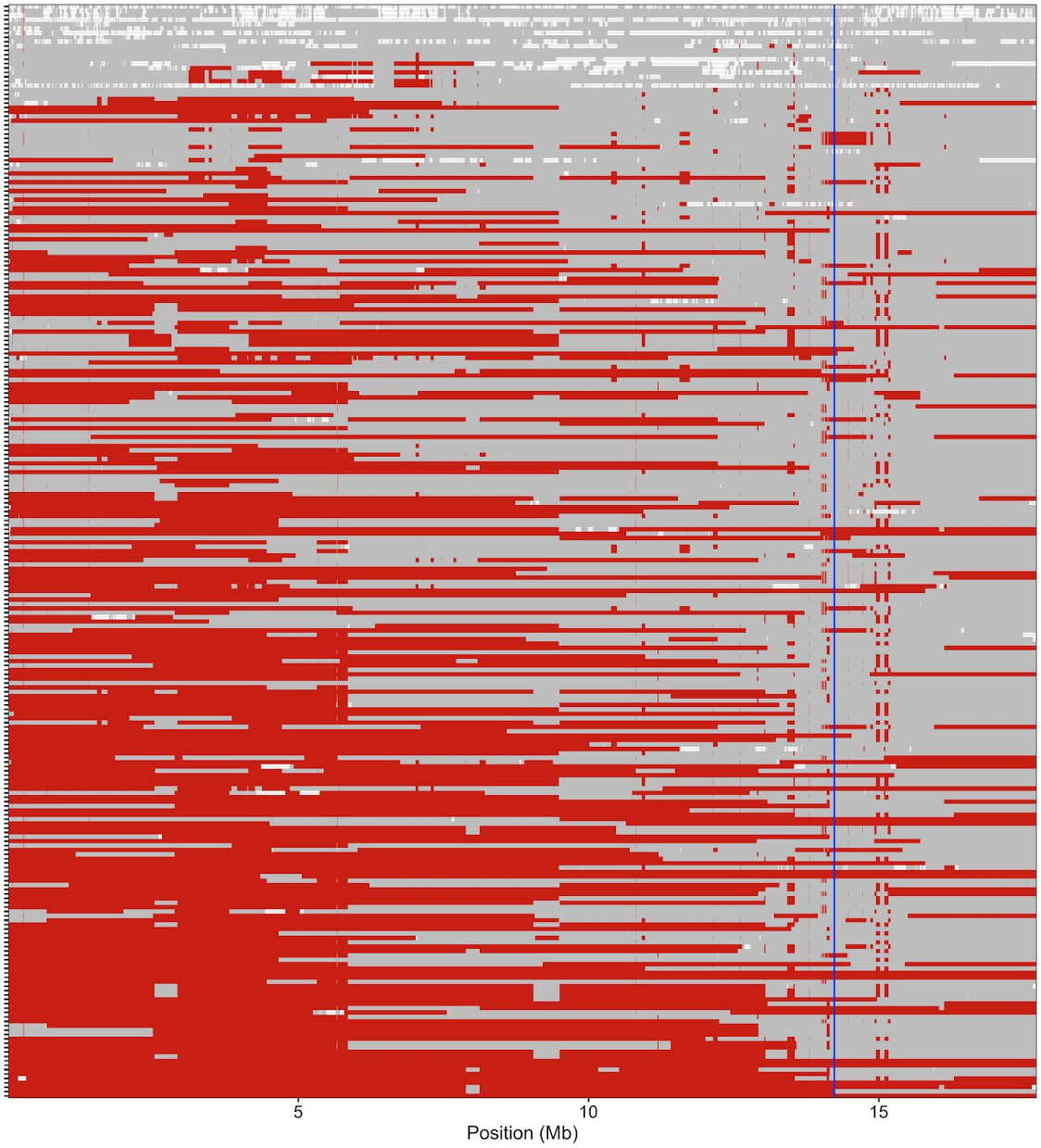

Fig. S9: *srg-37* is located far from the left arm of the X chromosome, which is frequently swept across 249 wild *C. elegans* strains

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.

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

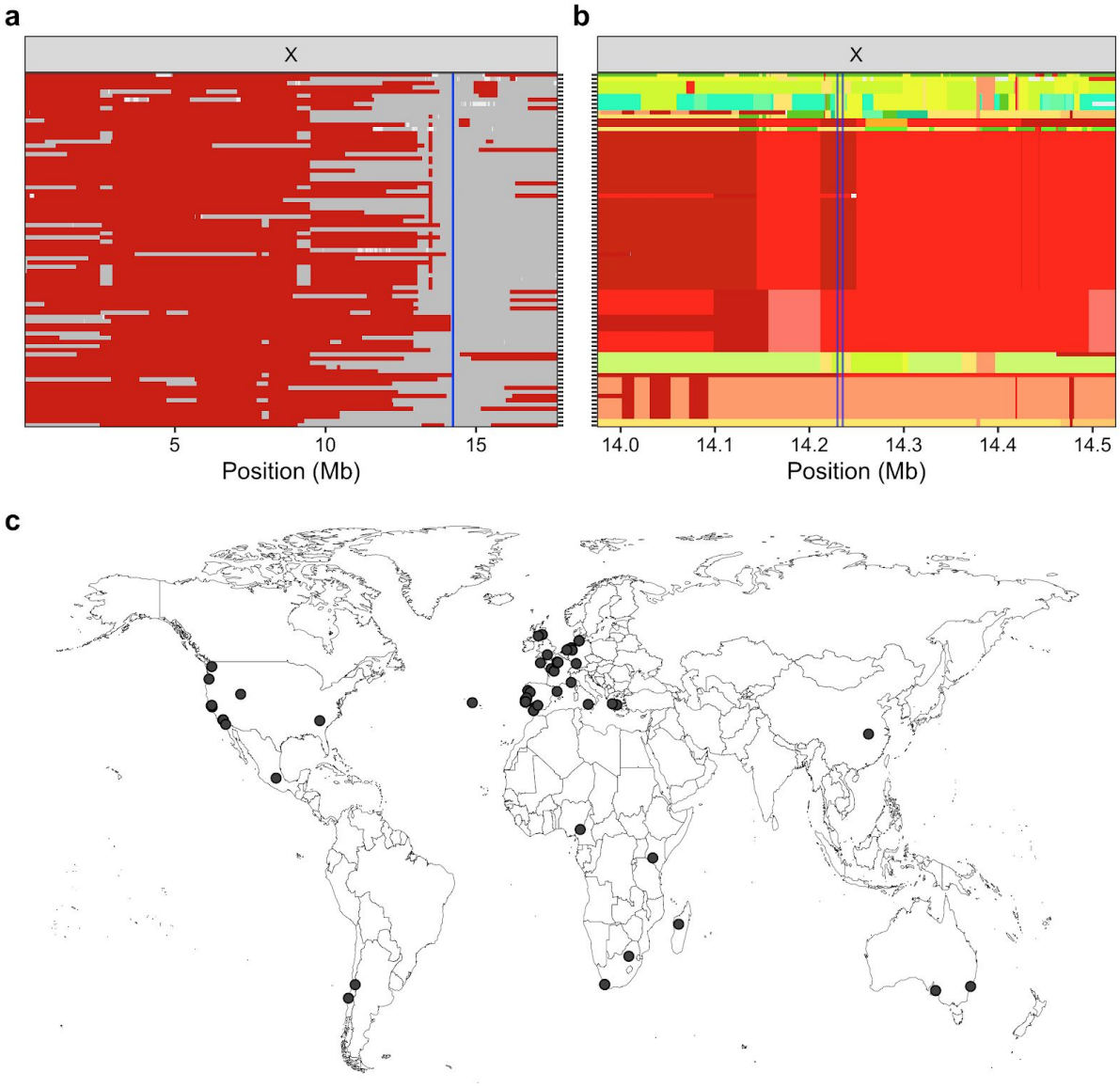

Fig. S10: *srg-37(ean179)* is outcrossed by multiple genotypes

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.

bioRxiv preprint doi: https://doi.org/10.1101/580803; this version posted March 18, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

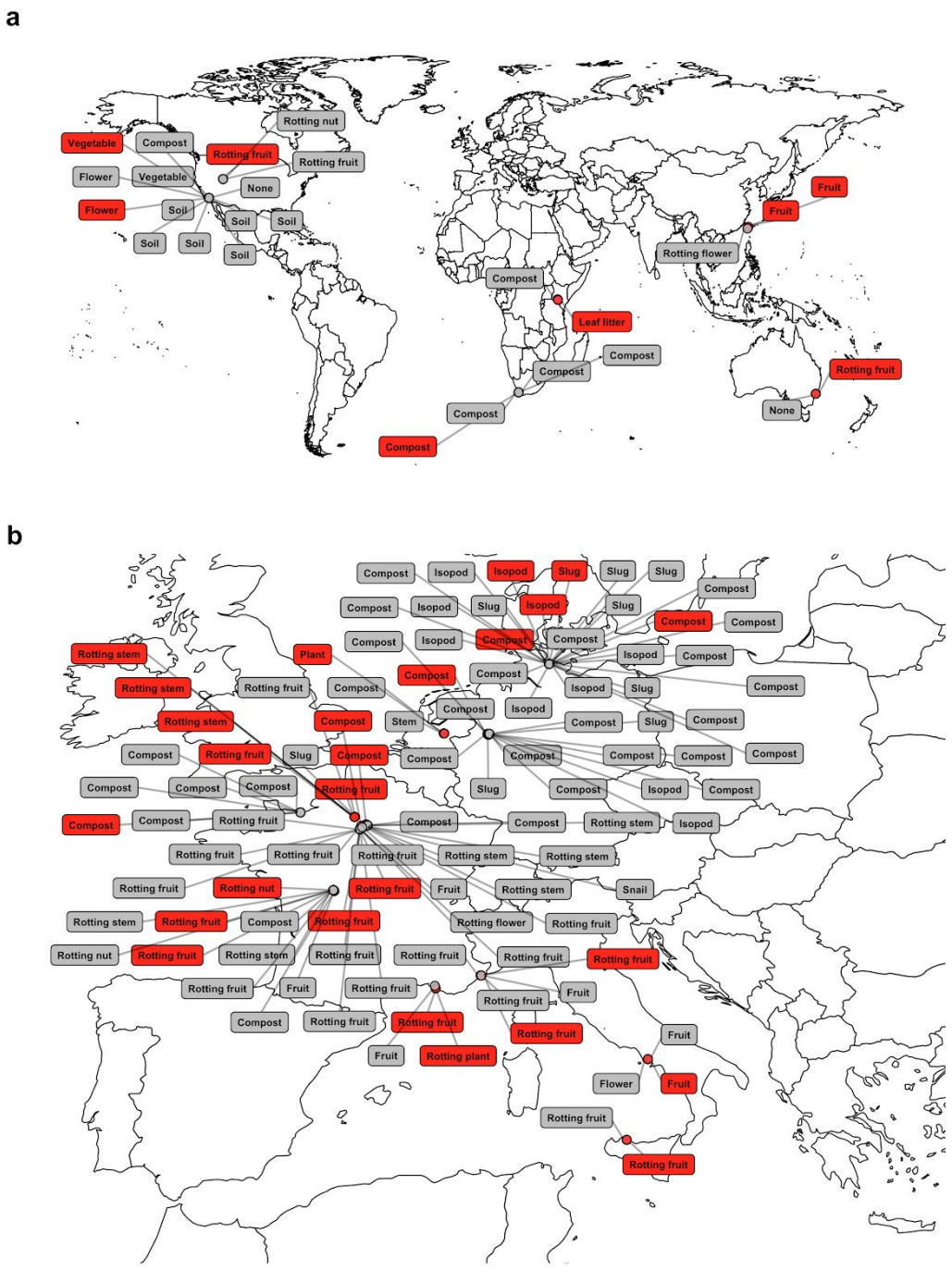

Fig. S11: Two *srg-37* genotypes are isolated in the local habitat across the world
Wild strains that are isolated from locations where both genotypes of *srg-37* were found are shown on the map of (a) world and (b) Europe. Each box indicates the substrate type where each strain is isolated, and is colored by the *srg-37* genotype of the isolated strain (grey: wild-type, red: *ean179*).