HHS Public Access

Author manuscript

Nat Ecol Evol. Author manuscript; available in PMC 2020 March 23.

Published in final edited form as:

Nat Ecol Evol. 2019 October ; 3(10): 1455–1463. doi:10.1038/s41559-019-0982-3.

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

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

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 natural genetic diversity affects the pheromone responses of individuals from diverse habitats. Here, we show that a niche-associated variation in pheromone receptor genes contributes to natural differences in pheromone responses. We identified putative loss-of-function deletions that impair duplicated pheromone receptor genes (*srg-36* and *srg-37*), which were shown previously to be lost in population-dense laboratory cultures. A common natural deletion in *srg-37* arose recently from a single ancestral population that spread throughout the world and underlies reduced pheromone sensitivity across the global *C. elegans* population. We found that many local populations harbor individuals with wild-type or a deletion allele of *srg-37*, suggesting that balancing selection has maintained the recent variation in this pheromone receptor gene. The two *srg-37* genotypes are

[*]Corresponding author: **Erik C. Andersen**, Associate Professor of Molecular Biosciences, Northwestern University, Evanston, IL 60208, USA, Tel: (847) 467-4382, Fax: (847) 491-4461, Erik.Andersen@Northwestern.edu.

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 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.

Competing interests

The authors declare no competing interests.

associated with niche diversity underlying boom-and-bust population dynamics. We hypothesize that human activities likely contributed to the gene flow and balancing selection of *srg-37* variation through facilitating migration of species and providing favorable niche for recently arose *srg-37* deletion.

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 signals act oppositely to promote either further reproductive growth or the development of a stress-resistant and long-lived dauer stage[2,3]. *C. elegans* secretes sugar-based pheromone compounds called ascarosides[4], and animals must measure the amount of remaining food and the ascaroside pheromones 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, dauer formation 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, most 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 interspecific 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 likely that the genetic controls of dauer formation are under natural selection. Although differences in dauer development among a small number of wild *C. elegans* strains have been described previously[10–15], no underlying natural genetic variant has been identified. Here, we integrate laboratory experiments, computational genomic analyses, and field research to further our understanding of the genetic basis underlying intraspecific variation in pheromone-mediated developmental plasticity. We identify natural genetic variation in responses to dauer pheromone and characterize a pheromone receptor allele that has spread around the globe.

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 dauer 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[16]. 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 three known dauer-inducing synthetic ascarosides (ascr#2, ascr#3, and ascr#5). We found significant variation in the dauer-pheromone responses among the strains tested, as measured by the fraction of individuals that enter the dauer stage (Fig. 1d, 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.

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)[17,18]. We found significant variation in the ascr#5 response with a broad-sense heritability estimate of 0.29 (H^2, SE=0.14) and a narrow-sense heritability estimate of 0.18 (h^2, SE=0.12) (Fig. 2a; Materials and methods). 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. Because

population structure can drive mapping of loci that are in interchromosomal linkage disequilibrium (LD) with causal QTL, we checked the LD among four QTL. We did not detect any obvious LD among these QTL (Supplementary Fig. 3), suggesting that multiple independent genomic loci underlie natural variation in the ascr#5 response.

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 (Supplementary Fig. 4), including the duplicated genes *srg-36* and *srg-37*, which encode ascr#5 receptors[19]. Both genes are expressed in the same pair of chemosensory neurons (ASI), which play an essential role in the dauer-pheromone response[20,21]. Notably, previous studies reported that both *srg-36* and *srg-37* are repeatedly deleted during long-term propagation of two independent laboratory-domesticated *C. elegans* lineages in high-density liquid cultures[19].

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)[22,23]. Although we could not find a large deletion that removes both *srg-36* and *srg-37*, we found only 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. 5). 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. 6). 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, Welch's t-test, $p = 3.152e-06$), suggesting that this deletion allele might cause a reduction in the acr#5 response.

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[24,25], 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. 7), 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[26]. 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, Supplementary Fig. 8). Differences in deletion allele frequencies between *srg-36* and *srg-37* suggest 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 isotypes (genome-wide genotypes) carry the 94-bp deletion allele, *srg-37(ean179)*.

Although *srg-36* and *srg-37* are duplicated genes that are activated by the same ligand and are 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[19]. 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. 5). 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 (Supplementary Fig. 9, Paired t-test, $p = 0.1981$; Materials and methods). It is more likely that differences in 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. 10)[27], 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 might allow *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 isotypes 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 (Supplementary Fig. 11, 12)[28]. Notably, we found that all 46 isotypes with the *srg-37* deletion exclusively share the swept haplotype at the *srg-37* locus (Fig. 4c). By contrast, none of 203 isotypes with wild-type *srg-37* carries the swept haplotype at the *srg-37* locus. 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 isotypes have an X chromosome that is swept more than 50% of its length but have diverse non-swept haplotypes at the *srg-37* locus (Supplementary Fig. 13). Additionally, the genome-wide tree of 249 wild C. *elegans* isotypes shows that the *srg-37* deletion is not present in many subpopulations (Fig. 4d). These results suggest that the *srg-37* deletion spread globally with the selective sweeps but has been purged after more recent outcrossing.

Two different *srg-37* genotypes coexist in local habitat and associate with different niches

These signatures of multiple outcrossing events imply the co-occurance of wild strains with and without the *srg-37* deletion in the same habitats. Indeed, we found that many local populations across the world harbor distinct individuals with either the wild-type *srg-37* or the deletion allele (Supplementary Fig. 14, see Materials and methods). Because each genotype can be adaptive to different environmental conditions, 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. 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[29]. By contrast, *C. elegans* were sampled predominantly in the dauer stage from animal and compost substrates[6,30]. We found that wild strains with the *srg-37* deletion were 67% enriched in rotting fruits (Fig. 4e, Supplementary Data 1, hypergeometric test, $p = 0.0026$). Thus, this allele is not only associated with lower dauer-pheromone responses but also with natural substrates that are known to support reproductive growth. Additionally, we analyzed F_{ST} statistics of the entire X chromosome for subpopulations from different substrates across shared geographic regions. Consistent with the niche association pattern of *srg-37* genotypes, we found the highest genetic divergence between the subpopulation from rotting fruit and the subpopulation from animal substrates at a genomic locus around the *srg-37* gene (Supplementary Fig. 15).

Discussion

Dauer pheromones are chemical signals that are perceived by sensory neurons using chemoreceptors and cGMP-mediated signaling[5,31]. 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)[32]. Genetic variation in the genes that mediate pheromone perception or downstream 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[33,34], mutations in these pathways might cause deleterious pleiotropic effects. Previous studies have shown that the ascr#5-receptors SRG-36 and SRG-37 were lost in two independent laboratory lineages of *C. elegans*[19], 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 only 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[19,35]. 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. The presence of three additional ascr#5-response QTL suggest that 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[19]. 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 neutrality statistics (Tajima's *D)* suggest that selection acts on these two genes differently. Our results indicate 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.* Substantial differences in amino acid sequences between SRG-36 and SRG-37 suggest that the SRG-37 protein is likely to have less ascr#5 binding affinity or weaker signal transduction activity than SRG-36. It is also possible that redundancy between SRG-36 and SRG-37 has been reduced since the time of gene duplication, and SRG-37 could gain sensitivities to other ascarosides while SRG-36 has maintained its ascr#5 specificity. Given the important role of the dauer stage in the long-term survival and dispersal of the species, purifying selection might act to conserve the primary ascr#5 receptor (SRG-36) in the *C. elegans* population to maintain the responsiveness to the dauer-inducing pheromone ascr#5.

In contrast to the rare deletion of *srg-36*, we identified a common deletion allele (18% allele frequency) of *srg-37* in the global *C. elegans* population. Notably, we discovered that strains harboring different *srg-37* genotypes (wild-type and deletion) have been found often in close proximity at various locations across the world, suggesting that balancing selection might have maintained both genotypes in local habitats. Previously, features of balancing selection

were also reported for a locus with other pheromone receptor genes (*srx-43* and *srx-44*) that underlie differences in *C. elegans* density-dependent foraging behavior[36,37]. Differences in food distribution can exert bidirectional fitness effects on foraging behavior. Similar to these effects, dauer formation can be disadvantageous during the population growth phase (boom phase) but beneficial during the dispersal phase (bust phase). Therefore, we hypothesize that the loss of *srg-37*, which reduces dauer formation, has trade-off effects between the boom and bust phases. Niche association patterns of *srg-37* genotypes support this hypothesis. We found that wild strains with the *srg-37* deletion are enriched in a rotting fruit niche, where ample bacterial food can support population growth during the boom phase. By contrast, the *srg-37* deletion is not enriched in wild strains isolated from animal carriers, which is consistent with known behavioral ecology during the bust phase when dauer larvae can readily hitchhike on other animals for their dispersal[6,8,9]. Our F_{ST} analysis also demonstrated significant genetic divergence at the *srg-37* locus between wild strains isolated from rotting fruit and animal carrier substrates. These observations suggest that the boom-and-bust population dynamics in wild habitats likely drive balancing selection of *srg-37*.

Population genomic analyses of the *srg-37* locus imply that the *srg-37* deletion arose recently and balancing selection might only have occurred for a short period of time. First, we found that strains with the *srg-37* deletion all share the same swept haplotype at the *srg-37* locus, which is estimated to have spread worldwide in the last few centuries[28]. Because mutation and recombination decrease linkage disequilibrium between a selected allele and the surrounding variants over time, this haplotype homogeneity suggests that the deletion allele arose recently. Second, we found no genomic signatures of long-term balancing selection. Tajima's D statistics for the *srg-37* locus did not show typical features of long-term balancing selection (D>>0). We also found that genetic diversity (π) is reduced at the *srg-37* locus in strains that carry the *srg-37* deletion versus strains that carry the *srg-37(+)* wild-type allele (Supplementary Fig. 16). This result is a signature of a recently established balanced situation[38]. We hypothesize that this recent balancing selection is related to human activities, which were also suggested to be drivers of the recent global selective sweeps[28]. Agriculture could have provided nutritious niches and therefore expanded boom phases spatiotemporally, which is likely to cause an increase in selective pressures to maintain the *srg-37* deletion. Furthermore, human migration could facilitate the worldwide gene flow of the *srg-37* deletion allele. Our studies imply that human civilization might exert a large impact on the natural selection and evolution of wild species.

Materials and methods

C. elegans strains

Animals were cultured at 20°C on modified nematode growth medium (NGMA) seeded with the *E. coli* strain OP50[39]. 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[22,23]. All strain information can be found in Supplementary Data 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[40] 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[41] dissolved in 0.4% ethanol or 0.4% ethanol alone. Animals were cultured for 52 hours at 25°C 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 7.28×10^8 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[42]. One cluster with lower fluorescence and 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).

Genome-wide association mapping

A genome-wide association (GWA) mapping was performed using phenotype data from 157 wild *C. elegans* strains. The dauer fraction of 157 wild strains in ascr#5-treated (800 nM) conditions were measured from four batches of experiments with three independent high-throughput dauer assays each. Contaminated, over-crowded ($n > 80$), or uncrowded ($n < 20$) samples were filtered out from the dataset. Normalized dauer fraction was calculated using a linear model, dauer fraction ~ batch. Genotype data were acquired from the latest VCF release (Release 20180527) from CeNDR that was imputed as described previously[22]. We used BCFtools[43] to filter variants that had any missing genotype calls and variants that were below 5% minor allele frequency. We used PLINK v1.9[44,45] to LD-prune the genotypes at a threshold of $r2 < 0.8$, 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[46]. 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[47]. The correlation matrix was generated using the *cor* function in the correlateR R package[48]. We set any eigenvalue greater than one from this analysis to one and summed all of the resulting eigenvalues[49]. 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.

Heritability calculations

Broad-sense heritability (H^2) and narrow-sense heritability (h^2) estimates were calculated using the phenotype data of 157 wild strains from the GWA mapping (ascr#5 800 nM). The *A.mat* and *E.mat* functions in the sommer R package were used to generate an additive genotype matrix and an epistatic genotype matrix, respectively, from the genotype matrix used for the GWA mapping[50]. These matrices were used to calculate the additive and epistatic variance components using the sommer *mmer* function. Variance components were used to estimate heritability and standard error through the *pin* function ($H^2 \sim$ V1 + V2 / V1 + V2 + V3, $h^2 \sim$ V1 / V1 + V2 + V3) in the sommer package.

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) with the following default parameters (t=1, k=19, w=100, d=100, r=1.5, c=10000, A=1, B=4, O=6, E=1, L=5, U=9, T=30, v=3). Optical/PCR duplicates were marked with PICARD (version 1.111)[22,51–53]. Alignments with greater than 100X coverage were subsampled to 100X using sambamba[54]. We called large deletions using the Manta structural variant caller (v1.4.0) using the default caller and filter settings (MinQUAL = 20, MinGQ = 15, MinSomaticScore = 30, MaxMQ0Frac = 0.4)[55].

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[24,25]. 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 95°C 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 *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). Then, 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 1.

Gene expression analysis of *srg-36* and *srg-37*

Gene expression levels of *srg-36* and *srg-37* at the L1 larval stage (WBls:0000024) in the N2 strain were analyzed from published whole-animal (WBbt:0007833) RNA-seq datasets (ERP003471, SRP000253, SRP000401, SRP003492, SRP003783, SRP008969, SRP010374, SRP034522, SRP040623, SRP058023)[56–65]. To equally weight each dataset with different number of replicates, mean values of FPKM for each dataset were used for gene expression comparisons.

Population genetics

Sliding window analysis of population genetic statistics (Tajima's D, F_{ST}, and π) was performed using the PopGenome package in R[66]. All sliding window analyses were performed using the imputed SNV VCF available on the CeNDR website with the most diverged strains XZ1516 set as the outgroup[22,67,68]. 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[69]. 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[70] with variants called by BCFtools[71] 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. To generate genome-wide tree, whole-population relatedness analysis was performed using RAxML-ng with the GTR+FO substitution model(DOI:10.5281/zenodo.593079). SNVs were LD-pruned using PLINK (v1.9) with the --indep-pairwise command `--indep-pairwise 50 1 0.95'. We used the vcf2phylip.py script (DOI:10.5281/zenodo.1257058) to convert the pruned VCF files to the PHYLIP format[72] required to run RAxML-ng. To construct the tree that included 249 strains, we used the GTR evolutionary model available in RAxML-ng[73,74]. Trees were visualized using the ggtree (v1.10.5) R package[75].

Substrate specificity analysis in the co-sampling zone

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 correlations between isolated substrate and the *srg-37* genotype of each isolate. Isolation of wild strains that share the 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[76].

Data and code availability

All data sets, including HTDA raw data, and code for generating figures are available on GitHub (https://github.com/AndersenLab/DauerSRG3637).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

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.

References

1. Cassada RC & Russell RL The dauerlarva, a post-embryonic developmental variant of the nematode Caenorhabditis elegans. Dev. Biol 46, 326–342 (1975). [PubMed: 1183723]
2. Golden JW & Riddle DL The Caenorhabditis elegans dauer larva: developmental effects of pheromone, food, and temperature. Dev. Biol 102, 368–378 (1984). [PubMed: 6706004]
3. Schaedel ON, Gerisch B, Antebi A & Sternberg PW Hormonal signal amplification mediates environmental conditions during development and controls an irreversible commitment to adulthood. PLoS Biol. 10, e1001306 (2012). [PubMed: 22505848]
4. Schroeder FC Modular assembly of primary metabolic building blocks: a chemical language in C. elegans. Chem. Biol 22, 7–16 (2015). [PubMed: 25484238]
5. Ludewig AH & Schroeder FC 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). [PubMed: 21093785]
7. Frézal L & Félix M-AC 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). [PubMed: 22081161]
9. Lee D et al. The genetic basis of natural variation in a phoretic behavior. Nat. Commun 8, 273 (2017). [PubMed: 28819099]
10. Viney ME, Gardner MP & Jackson JA Variation in Caenorhabditis elegans dauer larva formation. Dev. Growth Differ 45, 389–396 (2003). [PubMed: 12950280]
11. Harvey SC, Shorto A & Viney ME Quantitative genetic analysis of life-history traits of Caenorhabditis elegans in stressful environments. BMC Evol. Biol 8, 15 (2008). [PubMed: 18211672]
12. Green JWM, Snoek LB, Kammenga JE & Harvey SC Genetic mapping of variation in dauer larvae development in growing populations of Caenorhabditis elegans. Heredity 111, 306–313 (2013). [PubMed: 23715016]
13. Green JWM, Stastna JJ, Orbidans HE & Harvey SC Highly polygenic variation in environmental perception determines dauer larvae formation in growing populations of Caenorhabditis elegans. PLoS One 9, e112830 (2014). [PubMed: 25393108]
14. Diaz SA et al. Diverse and potentially manipulative signalling with ascarosides in the model nematode C. elegans. BMC Evol. Biol 14, 46 (2014). [PubMed: 24618411]
15. O'Donnell MP, Chao P-H, Kammenga JE & Sengupta P Rictor/TORC2 mediates gut-to-brain signaling in the regulation of phenotypic plasticity in C. elegans. PLoS Genet. 14, e1007213 (2018). [PubMed: 29415022]
16. 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). [PubMed: 27172224]
17. Evans KS et al. Correlations of Genotype with Climate Parameters Suggest Caenorhabditis elegans Niche Adaptations. G3 7, 289–298 (2017). [PubMed: 27866149]
18. Schulenburg H & Félix M-A The Natural Biotic Environment of Caenorhabditis elegans. Genetics 206, 55–86 (2017). [PubMed: 28476862]

19. McGrath PT et al. Parallel evolution of domesticated Caenorhabditis species targets pheromone receptor genes. Nature 477, 321–325 (2011). [PubMed: 21849976]
20. Bargmann CI & Horvitz HR Control of larval development by chemosensory neurons in Caenorhabditis elegans. Science 251, 1243–1246 (1991). [PubMed: 2006412]
21. Schackwitz WS, Inoue T & Thomas JH Chemosensory neurons function in parallel to mediate a pheromone response in C. elegans. Neuron 17, 719–728 (1996). [PubMed: 8893028]
22. Cook DE, Zdraljevic S, Roberts JP & Andersen EC CeNDR, the Caenorhabditis elegans natural diversity resource. Nucleic Acids Res. 45, D650–D657 (2017). [PubMed: 27701074]
23. Hahnel SR et al. Extreme allelic heterogeneity at a Caenorhabditis elegans beta-tubulin locus explains natural resistance to benzimidazoles. PLoS Pathog. 14, e1007226 (2018). [PubMed: 30372484]
24. Kim H et al. A co-CRISPR strategy for efficient genome editing in Caenorhabditis elegans. Genetics 197, 1069–1080 (2014). [PubMed: 24879462]
25. 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). [PubMed: 26187122]
26. Biswas S & Akey JM Genomic insights into positive selection. Trends Genet. 22, 437–446 (2006). [PubMed: 16808986]
27. 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).
28. Andersen EC et al. Chromosome-scale selective sweeps shape Caenorhabditis elegans genomic diversity. Nat. Genet 44, 285–290 (2012). [PubMed: 22286215]
29. Félix M-A & Duveau F Population dynamics and habitat sharing of natural populations of Caenorhabditis elegans and C. briggsae. BMC Biol. 10, 59 (2012). [PubMed: 22731941]
30. Barrière A & Félix M-A High local genetic diversity and low outcrossing rate in Caenorhabditis elegans natural populations. Curr. Biol 15, 1176–1184 (2005). [PubMed: 16005289]
31. Hu PJ Dauer. WormBook 1–19 (2007).
32. Lee SS & Schroeder FC Steroids as central regulators of organismal development and lifespan. PLoS Biol. 10, e1001307 (2012). [PubMed: 22505849]
33. Gumienny TL & Savage-Dunn C TGF-β signaling in C. elegans. WormBook 1–34 (2013).
34. Murphy CT & Hu PJ Insulin/insulin-like growth factor signaling in C. elegans. WormBook 1–43 (2013).
35. Gompel N & Prud'homme B The causes of repeated genetic evolution. Dev. Biol 332, 36–47 (2009). [PubMed: 19433086]
36. Greene JS et al. Balancing selection shapes density-dependent foraging behaviour. Nature 539, 254–258 (2016). [PubMed: 27799655]
37. Greene JS, Dobosiewicz M, Butcher RA, McGrath PT & Bargmann CI Regulatory changes in two chemoreceptor genes contribute to a Caenorhabditis elegans QTL for foraging behavior. Elife 5, (2016).
38. Charlesworth D Balancing selection and its effects on sequences in nearby genome regions. PLoS Genet. 2, e64 (2006). [PubMed: 16683038]
39. Andersen EC, Bloom JS, Gerke JP & 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). [PubMed: 24586193]
40. Boyd WA, Smith MV & Freedman JH Caenorhabditis elegans as a model in developmental toxicology. Methods Mol. Biol 889, 15–24 (2012). [PubMed: 22669657]
41. Zhang YK, Sanchez-Ayala MA, Sternberg PW, Srinivasan J & Schroeder FC Improved Synthesis for Modular Ascarosides Uncovers Biological Activity. Org. Lett 19, 2837–2840 (2017). [PubMed: 28513161]
42. 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).

43. 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). [PubMed: 21903627]
44. 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). [PubMed: 17701901]
45. Chang CC et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7 (2015). [PubMed: 25722852]
46. Endelman JB Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. Plant Genome 4, 250–255 (2011).
47. Qiu Y RSpectra. (Github).
48. Bilgrau AE correlateR. (Github, 2018).
49. Li J & Ji L Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity 95, 221–227 (2005). [PubMed: 16077740]
50. Covarrubias-Pazaran G Quantitative genetics using the Sommer package. R Found. Stat. Comput., Vienna https://cran.r-project.org/web/packages/sommer/vignettes/sommer.pdf (accessed 22 Mar. 2018) (2018).
51. Li H & Durbin R Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760 (2009). [PubMed: 19451168]
52. Picard Tools - By Broad Institute. Available at: http://broadinstitute.github.io/picard/.
53. Cook DE et al. The Genetic Basis of Natural Variation in Caenorhabditis elegans Telomere Length. Genetics 204, 371–383 (2016). [PubMed: 27449056]
54. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ & Prins P Sambamba: fast processing of NGS alignment formats. Bioinformatics 31, 2032–2034 (2015). [PubMed: 25697820]
55. Chen X et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 32, 1220–1222 (2016). [PubMed: 26647377]
56. Shin H et al. Transcriptome analysis for Caenorhabditis elegans based on novel expressed sequence tags. BMC Biol. 6, 30 (2008). [PubMed: 18611272]
57. Gerstein MB et al. Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science 330, 1775–1787 (2010). [PubMed: 21177976]
58. Mortazavi A et al. Scaffolding a Caenorhabditis nematode genome with RNA-seq. Genome Res. 20, 1740–1747 (2010). [PubMed: 20980554]
59. Stadler M & Fire A Wobble base-pairing slows in vivo translation elongation in metazoans. RNA 17, 2063–2073 (2011). [PubMed: 22045228]
60. Lamm AT, Stadler MR, Zhang H, Gent JI & Fire AZ Multimodal RNA-seq using single-strand, double-strand, and CircLigase-based capture yields a refined and extended description of the C. elegans transcriptome. Genome Res. 21, 265–275 (2011). [PubMed: 21177965]
61. Maxwell CS, Antoshechkin I, Kurhanewicz N, Belsky JA & Baugh LR Nutritional control of mRNA isoform expression during developmental arrest and recovery in C. elegans. Genome Res. 22, 1920–1929 (2012). [PubMed: 22539650]
62. Steijger T et al. Assessment of transcript reconstruction methods for RNA-seq. Nat. Methods 10, 1177–1184 (2013). [PubMed: 24185837]
63. Grün D et al. Conservation of mRNA and protein expression during development of C. elegans. Cell Rep. 6, 565–577 (2014). [PubMed: 24462290]
64. Wang J-J et al. The influences of PRG-1 on the expression of small RNAs and mRNAs. BMC Genomics 15, 321 (2014). [PubMed: 24884413]
65. Dillman AR et al. Comparative genomics of Steinernema reveals deeply conserved gene regulatory networks. Genome Biol. 16, 200 (2015). [PubMed: 26392177]
66. Pfeifer B, Wittelsbürger U, Ramos-Onsins SE & Lercher MJ PopGenome: an efficient Swiss army knife for population genomic analyses in R. Mol. Biol. Evol 31, 1929–1936 (2014). [PubMed: 24739305]
67. Browning BL & Browning SR Genotype Imputation with Millions of Reference Samples. Am. J. Hum. Genet 98, 116–126 (2016). [PubMed: 26748515]

68. Danecek P et al. The variant call format and VCFtools. Bioinformatics 27, 2156–2158 (2011). [PubMed: 21653522]
69. Warnes G, Gorjanc G, Leisch F, Man M. Genetics: Population genetics, R package version 1.3.6 (2012).
70. Browning BL & Browning SR Detecting identity by descent and estimating genotype error rates in sequence data. Am. J. Hum. Genet 93, 840–851 (2013). [PubMed: 24207118]
71. Danecek P, Schiffels S & Durbin R Multiallelic calling model in bcftools (-m).
72. Felsenstein J PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics 5, 164–166 (1989).

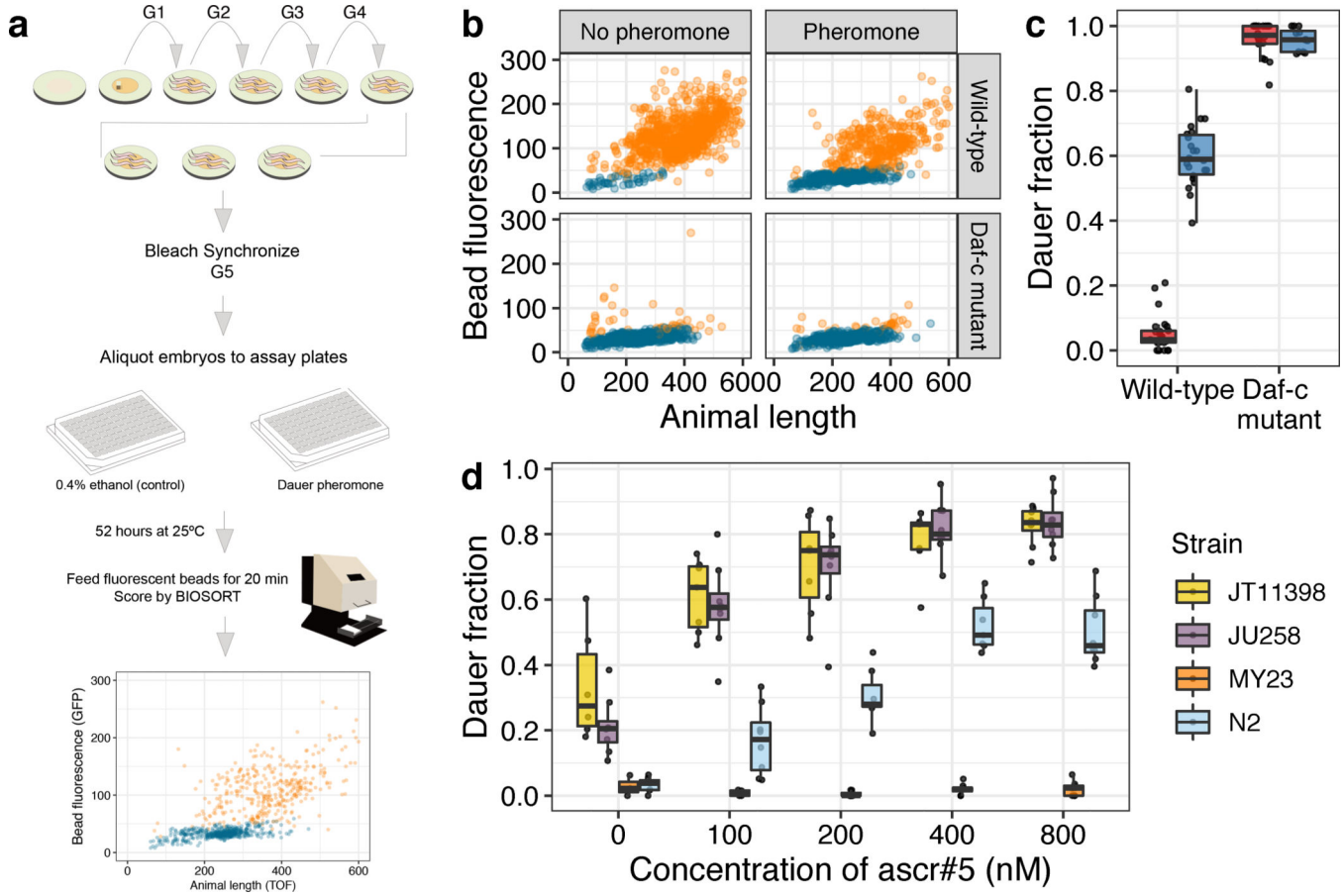

Fig. 1: A high-throughput dauer assay measures natural variation of dauer-pheromone response

(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 25°C 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 25°C for four divergent strains are shown with data points plotted behind. Concentrations of ascr#5 are shown on the x-axis, and fractions of dauer larvae are shown on the y-axis. (c, d) 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.

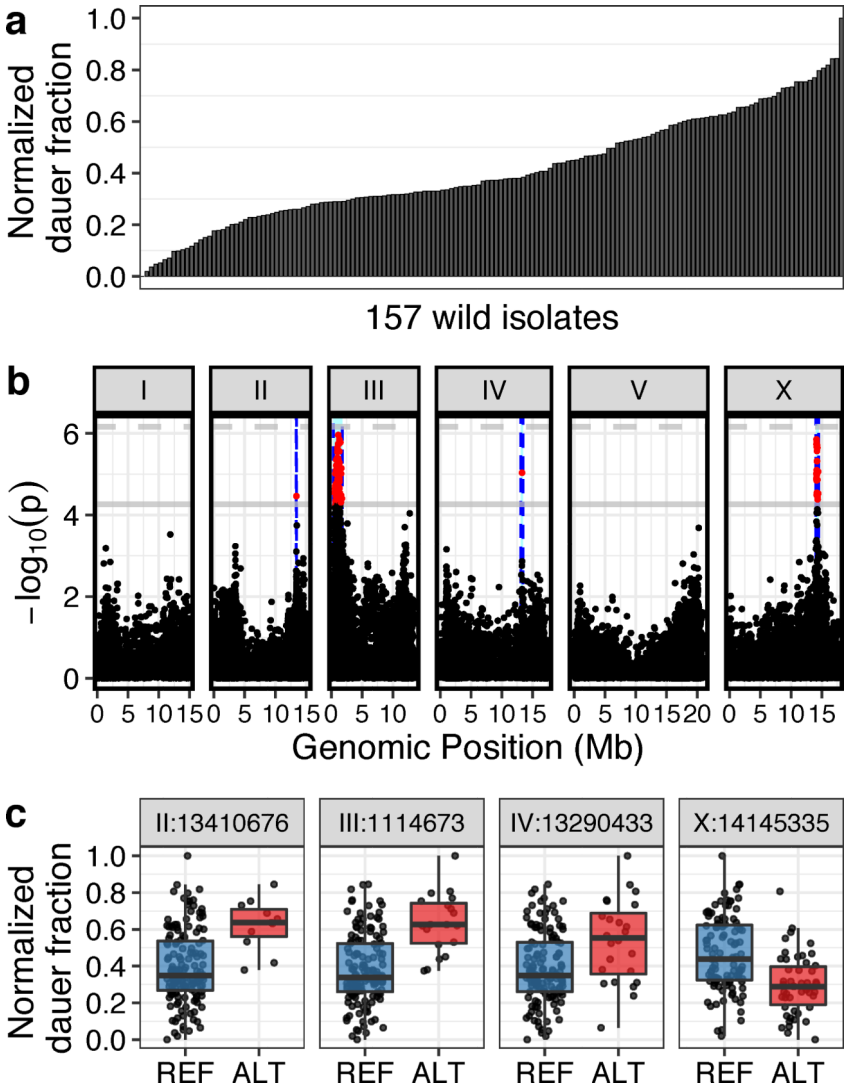

Fig. 2: Genome-wide association (GWA) mapping reveals four major loci underlying natural variation in dauer-pheromone response

(a) A bar plot for the natural variation of ascr#5-induced dauer formation at 25°C 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 157 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. Two significance thresholds are shown. The horizontal dashed line denotes the Bonferroni-corrected *p*-value threshold using all markers, and the gray horizontal line denotes the Bonferroni-corrected *p*-value threshold using independent markers correcting for linkage disequilibrium (genome-wide eigen-decomposition significance threshold). SNVs are colored red if they pass the second threshold. 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. 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.

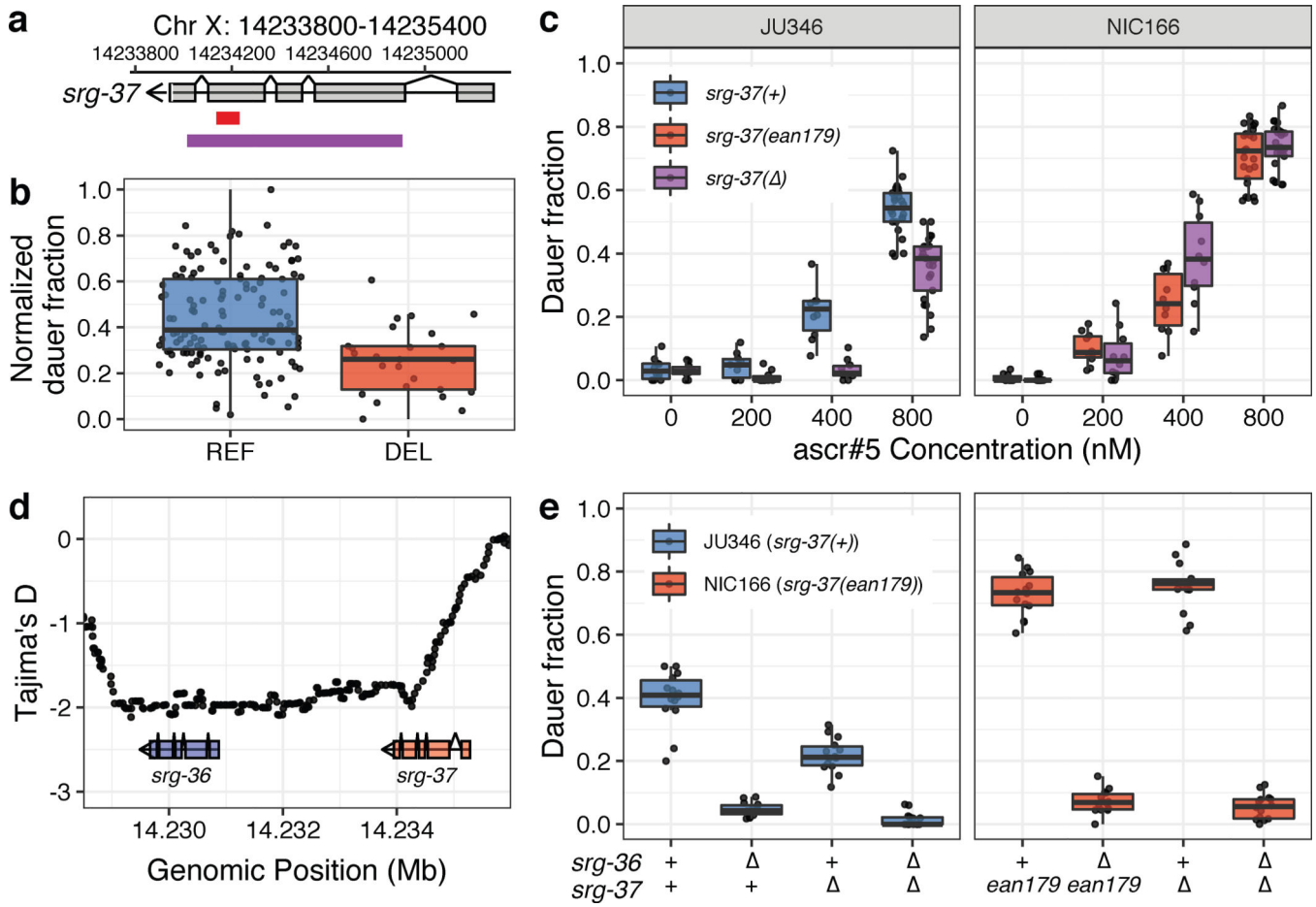

Fig. 3: Natural variant in the ascr#5 receptor gene, *srg-37*, underlies natural differences in dauer formation

(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 25°C 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. The gene structures of *srg-36* (blue) and *srg-37* (pink) are shown below the plot. The genomic position in Mb is plotted on the x-axis, and Tajima's D statistics are plotted on the y-axis. (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 25°C are shown with data points plotted behind. Genotypes of *srg-36* and *srg-37* are shown on the x-axis, where triangles represent the CRISPR-Cas9-mediated deletions. Fractions of dauer formation are shown on the y-axis. (b, c, e) 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.

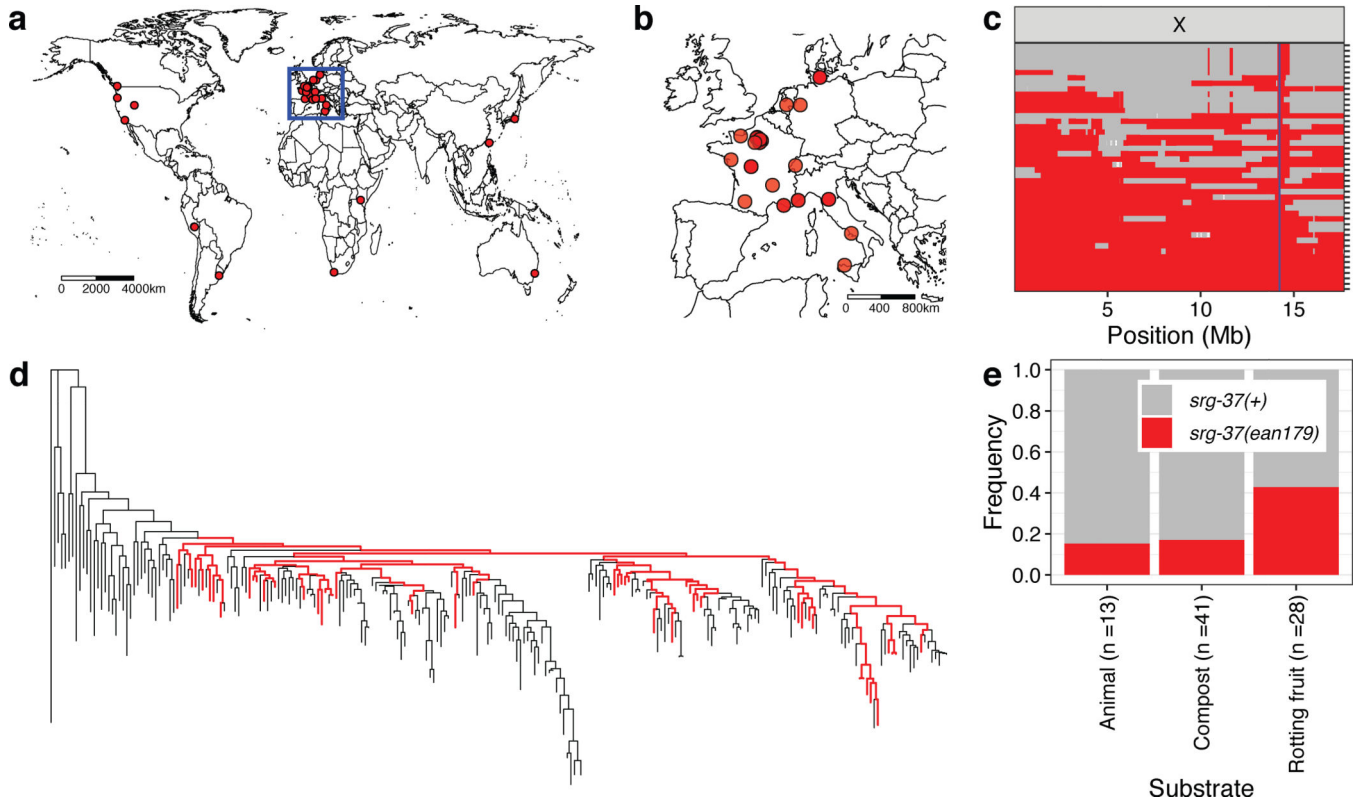

Fig. 4: Worldwide and niche-associated gene flow shape the ascaroside (ascr#5) pheromone receptor locus

(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. (a, b) Scale bars are shown in the map. (c) Sharing of the swept haplotype on the X chromosome among 46 wild isotypes with *srg-37(ean179)* is shown. Each row is one of the 46 isotypes, 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.