이학박사 학위논문

닉테이션 행동의 다양성에 대한 분자유전학적 연구

Genetic and molecular basis of the diversity of nictation behavior in *Caenorhabditis elegans*

2023 년 8 월

서울대학교 대학원

생명과학부

양 희 승

닉테이션 행동의 다양성에 대한 분자유전학적 연구

Genetic and molecular basis of the diversity of nictation behavior in *Caenorhabditis elegans*

지도교수 이 준 호

이 논문을 이학박사 학위논문으로 제출함
2023 년 6 월

서울대학교 대학원
생명과학부
양 희 승

양희승의 이학박사 학위논문을 인준함
2023 년 6 월

위 원 장 _____ (인)

부위원장 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

Genetic and molecular basis of the diversity of nictation behavior in *Caenorhabditis elegans*

A dissertation submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

To the Faculty of

School of Biological Sciences

at

Seoul National University

by

Heeseung Yang

June, 2023

Date approved

_____

_____

_____

_____

_____

_____

ABSTRACT

Genetic and molecular basis of the diversity of nictation behavior in *Caenorhabditis elegans*

Heeseung Yang
Dept. of Biological Sciences
The Graduate School
Seoul National University

Behavior is one of the most complex phenotypes of animals; numerous processes, from genetic programs to environments, regulate behaviors. Since the discovery of the laws of heredity and genetic material, how behavioral diversity arises from genetic variation has been one of the major questions in biology. In this study, I investigated the natural genetic variation underlying the diversity of the nictation behavior in the nematode *Caenorhabditis elegans*. Young *C. elegans* larvae grown under unfavorable conditions develop into dauer, an alternative developmental stage with many features for survival in harsh conditions. Dauer shows a stage-specific behavior, nictation, which helps its dispersal by promoting phoretic interaction. I performed nictation assays for 139 *C. elegans* wild strains worldwide and found differences in nictation fraction among them. Using genome-

wide association mapping and subsequent genetic analyses, I demonstrated that *nta-1*, a gene encoding an ortholog of a human hydroxysteroid dehydrogenase, regulates the diversity of nictation among wild strains. Experiments using transgenic animals identified sequence variation in the promoter region induces this difference by regulating the expression of *nta-1* in various tissues, especially in mesoderm-derived glial cells GLR. These findings provided a clear example of how sequence variation in the promoter region contributes to evolutionary novelty. Furthermore, the distribution of *nta-1* alleles varied across geographic regions, and population genetic analysis showed that the diversity of *nta-1* corresponds to ancient diversity in *C. elegans* populations that have been maintained by balancing selection. *nta-1* affected both nictation and post-dauer reproduction speed, suggesting that pleiotropy of *nta-1* on dispersal and reproduction influenced balancing selection of *nta-1* alleles.

Keyword : *C. elegans*, genetics, nictation, natural variation, behavioral diversity, GWAS

Student Number : 2015-20440

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Introduction

Behavioral diversity and genetics

Behavior is a significant characteristic of animals. It enables animals to respond to the environment, and ranges from simple reflex to complex patterns driven by multiple external and internal stimuli (Bendesky & Bargmann, 2011). Regulation of behavior involves numerous processes, including the development of organisms, memory and experience, environmental sensing, and decision-making. Differences between species and individuals can exist in all processes, creating behavioral diversity. Understanding behavioral diversity is crucial for comprehending individual differences, and one of the main objectives of genetics has been to clarify how genes influence behavior. However, characterizing specific behaviors at the molecular or cellular level is complex due to experimental challenges and the complexity of behaviors. Studies using simple model organisms have helped to understand behavior at the gene level. For example, Seymour Benzer is one of the pioneers in the behavioral genetics. Using forward genetics of *Drosophila*, he discovered various behavioral mutants including circadian rhythm mutants mapped to a single gene (Benzer, 1971; Konopka & Benzer, 1971).

C. elegans and dauer diapause

C. elegans is a small, transparent nematode worm widely used as an animal model organism for developmental and genetic research. It has many

advantages for genetic studies, including a short life cycle, ease of cultivation and maintenance, amenability to genetic manipulation, compact genome, and transparency. It was the first multicellular organism to have its genome sequenced and the first to have its brain's connectome revealed (Consortium, 1998; White et al., 1986), which makes it useful in the study of behaviors on the molecular and cellular level.

C. elegans has a life cycle of about 3.5 days when grown at 20°C. It hatches from an egg and passes through the larval stages L1, L2, L3, and L4 to reach the reproductive adult stage. However, when grown under unfavorable conditions such as high temperatures, high population densities, and food deprivation, L1 larvae can develop into L2d, followed by the diapause stage called dauers (Cassada & Russell, 1975; Golden & Riddle, 1984). Dauers have many distinct features compared to other stages, including their radially constricted body, thickened cuticles, increased stress resistance, and metabolic suppression, and can survive for several months without food intake (Cassada & Russell, 1975; Klass & Hirsh, 1976). These features make them advantageous for surviving and thriving under harsh conditions. Once dauers are in favorable conditions to grow, they resume development and can molt into L4. Interestingly, *C. elegans* dauers also exhibit a stage-specific behavior called nictation, which can help their dispersal (Lee et al., 2011).

Animal dispersal by phoretic interactions.

Dispersal is an important strategy used by organisms. It can help organisms survive, reproduce in uncertain environments, and promote organisms escape a region with limited resources. Sometimes, the interactions between species can influence the dispersal. Phoresy is a commensal interaction in which an organism uses another organism to move to another habitat, like hitchhiking (White et al., 2017). By providing a way for small organisms to overcome their limited mobility, phoresy allows them to increase the distance they can travel and the efficiency with which they can disperse. Back in the 19th century, Charles Darwin was intrigued by the phenomenon of less mobile small animals being found together at great distances, and suggested dispersal strategies of young freshwater shells hitchhiking on the feet or legs of birds, beetles and frogs (Darwin, 1859) (Darwin, 1882). Various phoretic interactions have been proposed since then, including the attachment of the ostracod to frogs (Lopez et al., 1999), the blister beetle larvae to bees (Saul-Gershenz & Millar, 2006), and mites to bird nostrils (Proctor & Owens, 2000).

Nictation behavior of nematodes

Nictation involves dauers standing on their tails and waving their bodies in three dimensions, increasing the probability that they will encounter and climb

onto the bodies of other organisms, such as isopods, slugs, and snails (Frézal & Felix, 2015). In nature, a *C. elegans* population has a boom-and-bust life cycle, with the population increasing rapidly when food is plentiful and decreasing rapidly when food is scarce (Felix & Braendle, 2010). Nictation can promote the dispersal of dauers developed in harsh environments. Once in a more favorable environment for development, dauers begin to feed and develop again, allowing the population to increase. *C. elegans* are usually collected in the dauer stage, suggesting that dauer is a dominant stage in their life cycle in the wild (Barriere & Felix, 2005). This implies that characteristics of dauers, such as nictation behavior, have greatly influenced the survival and evolution of nematodes.

Besides *C. elegans*, nictation behavior is conserved in many nematode species, including other *Caenorhabditis* species, such as *C. remanei, C. briggssae, C. japonica, and C. plicata*, and in other genera such as *Pristionchus pacificus* (Brown et al., 2011; Kiontke & Sudhaus, 2006). With the first observed nictation in these parasitic nematodes (Reed & Wallace, 1965), nictation also occurs in parasitic nematodes, where nictation is specific to the infective juvenile stage and is a parasitic behavior as it enables larvae to attach to hosts and infect them. Various species in *Steinernema, Strongyloides, Heterorhabditis, Nippostrongylus, Ancylostoma, Necator, Heligmosomoides*, and *Mermis* show nictation behavior

(Campbell & Gaugler, 1993; Castelletto et al., 2014; Gans & Burr, 1994; Granzer & Haas, 1991; Hernandez & Sukhdeo, 1995; Ishibashi & Kondo, 1990).

Previous research on nictation

It had been difficult to experimentally induce and quantify nictation, but the development of a nictation assay using a PDMS-based micro-dirt chip allowed genetic and molecular studies on nictation behavior (Lee et al., 2011). It was demonstrated that acetylcholine transmission in IL2 ciliated head neurons is essential for nictation (Lee et al., 2011) and that dauer-specific IL2 arborization regulates nictation (Schroeder et al., 2013). Another study showed that the pathways that control dauer development - Insulin/IGF-1 and TGF-β signaling pathways - also regulate nictation, with TGF-β signaling being epistatic to insulin signaling (D. Lee, H. Lee, et al., 2017). Neuropeptides such as *flp-7*, *flp-10*, *flp-11* and *flp-17* (Cockx et al., 2023; J. S. Lee et al., 2017) and piRNA (D. Lee, H. Yang, et al., 2017) were also found to play a role in regulating nictation. Transmission experiments with fruit flies and isopods revealed that nictation cloud facilitate dauer dispersal by carriers (D. Lee, H. Yang, et al., 2017; Lee et al., 2011).

Natural genetic variation to behavioral diversity

Studies that use a single or small group of reference strains of a model organism have an advantage in minimizing background effects and have contributed to the understanding of basic biological processes. (Gasch et al., 2016). For example, in the *C. elegans* field, most of the findings have relied on a lab-adapted strain N2 collected in Bristol. In nature, however, there are numerous natural genetic variations that are the raw materials of the evolutionary process. Natural variation is a large pool of mutations that shape the extensive diversity of our world. According to the 1000 genome project, there are average of 250-300 loss-of-function variants per person (Consortium, 2010). However, it is difficult to determine the function of genes that have already broken in each genetic background. Utilizing only specific genetic backgrounds can result in a limited comprehension of biological processes and phenotypic variety.

In quantitative genetics, the correlation between genotype and phenotype can be analyzed by several mapping methods using recombinant inbred lines (RIL), near-isogenic lines (NIL), and genome-wide association (GWA) to find quantitative trait locus (QTL) (Andersen & Rockman, 2022). QTL mappings using *C. elegans* have revealed genetic factors related to various behavioral traits such as gas response (McGrath et al., 2009), lawn-leaving behavior (Bendesky et al., 2011), foraging behavior (Greene, Brown, et al., 2016; Greene, Dobosiewicz, et al.,

2016) and plasticity on CO2 response (Beets et al., 2020). In addition, a study on nictation using RIL demonstrated the significance of piRNA-rich regions, though it did not identify a specific gene (D. Lee, H. Yang, et al., 2017).

Purpose of Research

Nictation is a stage-specific behavior of *C. elegans* that is predicted to be important in their survival. While previous studies have revealed several genetic factors that regulate this behavior, the variations of specific genes that induce diversity in nictation remain to be elusive. In this study, using more than 100 wild *C. elegans* strains from around the world, I investigated the natural variation in nictation and its underlying causative genetic variants using GWA mapping techniques that have recently become available in *C. elegans*. This study will contribute to our understanding of how genetic variation can cause behavioral diversity and generate evolutionary significance.

Materials and Methods

Strains used in this study.

Wild strains. 139 *C. elegans* wild strains were obtained from *C. elegans* Natural Diversity Resource (CeNDR).

AB1, AB4, BRC20067, CB4852, CB4854, CB4856, CB4932, CX11259, CX11262, CX11264, CX11271, CX11276, CX11285, CX11307, CX11314, CX11315, DL200, DL226, DL238, ECA246, ECA248, ECA259, ED3005, ED3011, ED3012, ED3017, ED3040, ED3046, ED3052, ED3073, EG4347, EG4724, EG4725, EG4946, GXW1, JT11398, JU1088, JU1172, JU1200, JU1212, JU1213, JU1242, JU1395, JU1409, JU1440, JU1491, JU1530, JU1568, JU1580, JU1581, JU1586, JU1652, JU1896, JU2001, JU2007, JU2316, JU2464, JU2466, JU2513, JU2519, JU2522, JU2526, JU258, JU310, JU323, JU346, JU360, JU367, JU393, JU397, JU406, JU440, JU561, JU642, JU751, JU774, JU775, JU778, JU782, JU792, JU830, JU847, KR314, LKC34, MY1, MY10, MY16, MY18, MY23, N2, NIC1, NIC166, NIC195, NIC199, NIC2, NIC207, NIC231, NIC236, NIC242, NIC251, NIC252, NIC255, NIC256, NIC258-262, NIC265-269, NIC271, NIC272, NIC274-277, NIC3, PB303, PS2025, PX179, QG536, QG556, QG557, QW947, QX1212, QX1233, QX1791-1794, RC301, WN2001, WN2002

NILs.

LJ1301 *ysIR30*(*nict-2* QTL, CB4856 > MY23),

LJ1317 *ysIR31*(*nict-2* QTL, MY23 > CB4856),

LJ1318 *ysIR32*(*nict-2* QTL, MY23 > CB4856),

LJ1319 *ysIR33*(*nict-2* QTL, MY23 > CB4856),

LJ1320 *ysIR34*(*nict-2* QTL, MY23 > CB4856)

Mutants.

rrf-3(pk1426) for RNAi experiments

LJ1321 (*nta-1(ys101)* in CB4856), LJ1322 (*nta-1(ys102)* in LJ1318), LJ1323 (*nta-1(ys103) in* N2)

Transgenic strains.

N2;*Ex[nta-1p*^{CB4856}::gfp, rol-6(su1006)]*,

N2;*Ex[nta-1p*^{MY23}::gfp, rol-6(su1006)]*,

N2;*Ex[nta-1p^{CB4856}::gfp]*, N2;*Ex[nta-1p^{MY23}::gfp]*,

N2;*Ex[nta-1p^{CB4856}::gfp, nta-1p^{MY23}::mcherry]* ,

N2;*Ex[nta-1p ^{CB4856}::gfp, egl-3p::mcherry]*,

N2;*Ex[nta-1p^{CB4856}::gfp, nep-2p::mcherry]*,

LJ1321;*Ex[nta-1p^{CB4856}::nta-1 ^{CB4856}::sl2::gfp]*,

LJ1321;*Ex[nta-1p^{CB4856}::nta-1 ^{MY23}::sl2::gfp]*

LJ1321;*Ex[nta-1p^{MY23}::nta-1 ^{CB4856}::sl2::gfp]*,

LJ1321;*Ex[nep-2p::nta-1 ^{CB4856}::sl2::gfp]*,

LJ1321;*Ex[egl-6p::nta-1 ^{CB4856}::sl2::gfp]*

*The asterisked *nta-1p* is about 2 kb, and the rest are about 850 bp.

***C. elegans* culture.**

Worms were grown at 15 or 20°C under standard condition (Brenner, 1974), except for dauer induction.

Dauer induction.

About 15 young adults were transferred to the pheromone plate seeded with *Escherichia coli* OP50 and cultured at 25 °C to lay eggs. Pheromone plates contain agar (10 g/L), agarose (7 g/L), NaCl (2 g/L), KH2PO4 (3 g/L), K2HPO4 (0.5 g/L), cholesterol (8 mg/L) and synthetic pheromone-ascarosides C7 (daumone 1, ascaroside 3), C6 (daumone 2, ascaroside 1), C9 (daumone 3, ascaroside 2) (2 mg/L each) (Butcher et al., 2007; Jeong et al., 2005). After four days, dauers among progenies were identified by their constricted thin bodies.

Nictation Assays.

3.5% agar solution was poured onto a poly-dimethylsiloxane (PDMS) mold to prepare micro-dirt chips for the nictation assay. A microwave was used to dissolve the agar powder. After the agar was solidified, the agar was removed from the PDMS mold and dried at 37 °C for 90 minutes. Dauers on the pheromone plates were collected by glass capillary (Kimble Chase Life Science and Research Products LLC.) using M9 buffer and transferred onto the dried micro-dirt chip.

After 30-40 min, when most of the dauers were moving, the proportion of the nictating dauers among the moving dauers was measured. For each biological replicate, the nictation fraction was measured three times in a row. The mean of these three technical replicates represented the nictation fraction of a biological replicate. All nictation assays were performed in a closed room with a thermo-hygrostat set to maintain a temperature of 25 °C and a humidity of 30%. One hundred thirty-eight wild isolates were tested with CB4856 control, and their nictation fractions were normalized to that of CB4856. In all the other nictation assays, nictation fractions were regressed by fitting a linear model, nictation fraction ~ assay trial, and the total mean of raw nictation fraction was added to each residual.

Calculation of narrow-sense heritability

Nictation fraction of 137 wild isolates were used to calculate an estimate of narrow-sense heritability (h^2). The mmer function in the sommer R package was used to calculate the additive variance. This was used to calculate the estimate of h^2 using pin function in the sommer package (h^2~ V1 / V1+V2).

Genome-wide association mapping

Before mapping, nictation fractions of 137 wild isolates were scaled to have a mean of zero with a standard deviation of one. Genome-wide association mapping and fine mapping of QTL were performed in CeNDR using NemaScan as described previously (Widmayer et al., 2022), using genotype data of the latest VCF release (Release 20220216) from CeNDR (Cook et al., 2017; Widmayer et al., 2022). Hard-filter VCF and imputed VCF were used for GWA mapping and fine mapping, respectively. In the GWA mapping, regions of interests were extended as interval of ± 150 SNVs from the significant marker above the eigen-decomposition significance threshold, defining a QTL. Region of interests within 1000 SNVs were grouped as a single QTL.

Generation of NILs.

All NILs were generated from CB4856 and MY23 parents. CB4856 was backcrossed to the MY23 parents ten times while selecting the CB4856 genome near the QTL interval on chromosome II by checking the single-nucleotide polymorphism (SNP) of II:3,768,531 and II:4,286,410. To genotype 3,768,531, I used EcoRV restriction fragment length polymorphism (RFLP) of single-worm PCR products amplified by 5'-CGGTTCGGCTGGAAATTTGG-3' and 5'-CAGT TGCTC CCGTTAAAGCG-3' primers. To genotype 4,286,410, I used HaeIII

RFLP of single-worm PCR products amplified by 5'-TCGTAAATCCACACCT GCGG-3' and 5'-CACGTGGAAAACCGCATCTG-3' primers. In the same way, MY23 was backcrossed to the CB4856 parents. Genomes of NILs were analyzed by short-read sequencing. The remaining unwanted genomes in chromosome II were removed by further backcrossing to generate LJ1301 and LJ1317. LJ1318, LJ1319, and LJ1320 were generated by further backcrossing of LJ1317.

RNAi nictation assay.

C. elegans RNAi feeding library from Julie Ahringer's group was used. Each target cell in the library was streaked and cultured in LB broth containing ampicillin. RNAi plates and RNAi pheromone plates containing 1mM IPTG were prepared. *rrf-3*(pk1426) mutant was used. To avoid OP50 being transferred together, I transferred worms three times. L4 worms on the OP50 plates were transferred to empty fresh plates without bacteria and allowed to move for 30 min. Worms were then transferred to each RNAi plate. After 30 minutes, Worms were transferred to new RNAi plates and incubated to lay eggs. After three days, about 15 young adults of F1 progenies were transferred to the RNAi pheromone plates. After another four days, dauers were collected, and their nictation fractions were measured as described above.

CRISPR/Cas9-mediated mutagenesis.

To induce *nta-1* deletion, I designed two sgRNA sequences 5' - ACCTTCTCAAGAGCACCTGGTGG – 3' and 5' – GTGATCTATGTTGTGCCT GGCGG - 3' that together can induce deletion of most of the *nta-1* sequence. pJL3021 and pJL3022, each containing one of these two sgRNAs, were derived from pJW1219 CRISPR/Cas9 plasmid containing sgRNA(F+E), edited by using the Q5 site-directed mutagenesis kit (E0554; New England Biolabs). CRISPR-edited worms were identified by co-CRISPR using PJA50 plasmid and AF-JA-76 repair template that induce *unc-58(e665)* (Arribere et al., 2014). The injection mixture with a final concentration of 30 ng/ul for pJA50, 20 ng/ul for *unc-58* repair template, and 50 ng/ul for both pJL3021 and pJL3022 was injected into young adults of the target background strains. F1 progenies with unc phenotype were singled out to a new plate and incubated to lay enough eggs. After they lay eggs, F1 progenies are genotyped after single-worm lysis. Genotyping was accomplished by determining the length of the PCR product, including the *nta-1* deletion site, amplified using 5'- CTGTTCGTGGGATTGTTTGA - 3' and 5'- GGCTCAACCCGTAGAAATTG -3' primers. F2 progenies were genotyped in the same way, and the F2 progenies with homozygous *nta-1* mutation were backcrossed to each background strain for at least three generations.

F1 hemizygous nictation assay.

25 to 30 of day one adult CB4856 or LJ1321 hermaphrodites were cultured for 3-4 days to lay eggs. Worms were transferred to a new plate each day to check their egg-laying. After 3-4 days, when their egg-laying had almost stopped, they were mated with LJ1318 or LJ1322 males. The next day, eighteen of the mated hermaphrodites were transferred to a pheromone plate. The mated hermaphrodites were distinguished by the formation of copulatory plugs around the vulva. Four days later, the nictation fraction of F1 dauers was measured. After the assays were completed, the dauers on the nictation chips were collected again and transferred to the new fresh plate with OP50 food. The next day, the male-to-hermaphrodite ratio of the recovered dauers was counted to validate that the measured dauers represented the pool of F1 progenies.

Generation of the transgene construct.

To generate all *promoter::gfp* or *promoter::mCherry* plasmids, each promoter PCR template was inserted into the GFP vector pPD95.77 or the mCherry vector pPD117.01. All plasmids for rescue experiments (*promoter::nta-1::sl2::gfp*) were generated by modification of the pEM1 vector (Macosko et al., 2009). All constructs were made using the Gibson assembly cloning kit (E5510; New England Biolabs).

Generation of transgenic animals.

The plasmid constructs were injected into the gonad of the young adult animals as previously described (Mello et al., 1991). The injection mixture contained plasmids for fluorescent protein expression at 100 ng/μL, and/or plasmids for *nta-1* rescue at 50 ng/μL, and/or plasmids for *rol-6(su1006)* transgenic marker at 100 ng/μL.

Protein structure prediction.

Predictions were performed using Colabfold 1.5.2. using CDS of *nta-1* of CB4856 and MY23 as query sequences (Mirdita et al., 2022). The structures of the proteins were visualized using UCSF chimera X (Pettersen et al., 2004).

Microscopy

ZEISS LSM700 confocal microscope and ZEN software (Carl Zeiss) were used to obtain images of fluorescence protein expression. When using a confocal microscope, animals were fixed on the 3% agar pad on the slide glass and paralyzed with 3 mM levamisole. Leica M205 FA fluorescence stereo microscope was used to obtain images of dumpy-like dauers.

Population genetics.

A VCF file containing 58 biallelic SNVs from 550 wild *C. elegans* genomes (Cook et al., 2017) was converted to the PHYLIP format to generate *nta-1* promoter tree. The distance matrix and pseudo-rooted (ECA2199) neighbor-joining tree were made from this PHYLIP file using dist.ml and the NJ function using the phangorn (version 2.11.1) R package. The tree was visualized using the ggtree (version 3.6.2) R package. Using this tree, two major haplotypes at the *nta-1* promoter locus were classified. To visualize the geographic distribution of *nta-1* haplotypes, the latest strain data from the CeNDR (Release 20220216) was used to get collection information of each isolate. A genome-wide neighbor-joining tree for 550 wild isotypes was obtained from the CeNDR (Cook et al., 2017). The tree was pseudo-rooted with ECA2199 and visualized using the ggtree (version 3.6.2) R package. Tajima's D statistics for *nta-1* locus were obtained from Lee et al. 2021 (Lee et al., 2021), where Tajima's D was calculated from the allele frequency spectrum of SNVs within 10 kb window across 328 wild genomes.

Measurement of total brood size.

To obtain synchronized embryos, about thirty of day 1 adults were transferred onto fresh NGM plates seeded with OP50 to lay eggs at 20°C. After an hour, adult worms were fully removed from the plate. After 2 days, the

synchronized L3-L4 worms were individually transferred to different plates. After reaching adulthood and starting to lay eggs, the adult worms were transferred to a new plate every day for two days. The number of progenies in each plate was counted after two days from transfer and summed to obtain the total brood size of an individual worm. Brood sizes were normalized by fitting a linear regression, brood size ~ test date.

Initial brood size after dauer recovery

Thirty dauers were transferred onto fresh NGM plates seeded with OP50. Dauer tended to move out of the OP50 lawns, so I checked on them for about 30 minutes and transferred them back into the OP50 lawn. After 40 hours, the worms were transferred to new plates, and again after 8 hours, the worms were removed. The number of worms remaining was counted at 40 and 48 hours. After two days, the number of progenies in each plate was counted. The number of offspring on the first plate was divided by the number of animals remaining after 40 hours, and the number of offspring on the second plate was divided by the number of animals remaining after 48 hours, and the two were added together to calculate the mean brood size up to 48 hours after dauer transfer. Brood sizes were normalized by fitting a linear regression, brood size ~ test date.

Initial brood size from eggs

Synchronized embryos were obtained as described above (Measurement of total brood size.**)**. After 2 days, 30 of synchronized L3-L4 worms were transferred onto a new plate. The next day, 70 hours from the synchronized worms were laid from their parents, the number of remaining worms was counted, and the worms were removed from the plate. The number of progenies in the plate was counted after two days. Brood sizes were normalized by fitting a linear regression, brood size ~ test date.

Results

Identification of a QTL for nictation by GWA mapping.

Wild isolates of *C. elegans* have many natural genetic variations. I speculated that nictation behavior is quantitatively different among wild isolates, and the underlying major genetic locus can be discovered in genome-wide association mapping. To determine if there are differences between wild isolates in nictation behavior, I induced dauer and measured the nictation fraction of 139 wild isolates from around the world (Figure 1, Figure 2). All the strains tested showed nictation behavior, and the nictation fractions of them varied over a wide range (Figure 3). An estimate of narrow-sense heritability (h^2), the proportion of phenotypic variance attributable to additive genetic variance, was 0.41. Two strains, JU310 and CB4852, were excluded from the analysis as dauers of two strains were shorter and plumper compared to the other strains (Figure 4). They had resistance to 1% SDS like normal N2 dauers, but both had very low nictation fractions, presumably due to their morphological features. Their morphological features could dominantly suppress their nictation over other genetic factors.

Next, I performed genome-wide association mapping of nictation to find the molecular basis for the natural variation of the nictation fraction. GWA was performed in CeNDR using NemsScan (Cook et al., 2017; Widmayer et al., 2022). GWA mapping revealed 4 significant single-nucleotide variants (SNV), with a peak SNV at 4,240,841 which explained 24.7 % of the phenotypic variance (Figure

5). Four significant SNVs spanned hundreds-kb range, which defined a *nict-2* QTL that spans from 3,732,152 to 4,766,883 on chromosome II. The strains divided into two groups based on the peak SNV since the SNVs that cloud divide all measured strains into two groups were used in the GWA mapping. The alternative (ALT) group had a higher nictation fraction than the reference (REF) group, which shares the same peak SNV with the reference strain N2 (Figure 6). Among 137 strains, 13 strains had an alternative peak SNV.

Fine mapping using NILs narrowed *nict-2* to 90 kb interval.

GWA mapping suggests genetic variants associated with the trait based on the linkage disequilibrium but uses a subset of SNV in the analysis. In addition, there are many other variants in the genome besides SNVs, and generally, fine mapping is needed to find the actual causal variation. At first, to validate the QTL effect, introgression lines for the chromosome II QTL were generated. CB4856 and MY23 were used as representative strains in the REF and ALT groups, respectively. (Figure 6b). CB4856 was selected because it was an intensively studied strain, and so much information was available. MY23 was, interestingly, the only isotype found in three different regions, and several other strains in the ALT group had unhealthy issues. After backcrossing 10 times, most of the genomic regions were replaced except for the target QTL interval. The nictation fractions of

these near-isogenic lines (NILs) validated the effect of the *nict-2* locus (Figure 7). The MY23 genome around the QTL position increased the nictation in the CB4856 genetic background, whereas the CB4856 genome decreased the nictation in the MY23 genetic background.

Next, I tried to narrow the QTL interval using near-isogenic lines (NILs) with a short MY23 genome in the CB4856 background. Several other NILs were generated by backcrossing LJ1317 with the background strain CB4856 for fine mapping of this QTL. Comparing their nictation fractions, I found that the QTL effect of LJ1317 was split into two loci of about 660 kb and 90 kb in length, respectively. (Figure 8). LJ1318 strain retained a significant QTL effect, whereas LJ1319 strain, which has a 90-kb shorter MY23 genome, did not show the QTL effect. Since this 90 kb region was in the first QTL interval found in the GWA mapping, I focused on this 90 kb QTL.

***nta-1* in the *nict-2* locus regulates nictation.**

The 90 kb *nict-2* locus (3,708,886 – 3,799,634) contained 31 protein-coding genes, 5 pseudogenes, 2 ncRNAs, and one 21U-RNA gene (Figure 9). Among these genes, 19 protein-coding genes had valid variants between CB4856 and MY23 (Table 1). Based on their known expression profiles and predicted effects of the variants, I selected primary candidate genes and performed RNAi

knockdown experiments. Knockdown of none but *F12E12.11(F12E12.c)* and *F12E12.11 + F12E12.1*(*F12E12.i*) affected nictation, indicating that *F12E12.11* is the likely causative gene for the *nict-2* QTL effect (Figure 10). In addition, *F12E12.11* had been reported in a previous study to show increased expression under dauer-inducing conditions and differential expression in two strains N2 and DR1350 (Harvey et al., 2009).

To directly validate the function of *F12E12.11* in nictation, *F12E12.11* deletion mutants of CB4856 and LJ1318 NIL were generated using CRISPR/Cas9 (Figure 11). While *F12E12.11* deletion in CB4856 increased the nictation fraction, deletion in LJ1318 NIL had no effect. Since CB4856 shares similar sequences with reference strain N2 around the *F12E12.11* (Figure 12), *F12E12.11* deletion mutation was introduced in the N2 genetic background to check whether it induces the same effect on nictation. As expectedly, the *F12E12.11* deletion in N2 also increased the nictation fraction (Figure 11). These results suggest that *F12E12.11* in CB4856 and N2 reduces nictation, whereas *F12E12.11* in MY23 does not function in regulating nictation. Based on these results, *F12E12.11* was named *nta-1* (*nta*; NicTation Altered).

The nictation phenotypes of *nta-1* mutants suggested *nta-1* would be the causal gene that generated the QTL effect of *nict-2* locus. To confirm this, I performed a reciprocal hemizygosity F1 test (Figure 13). The reciprocal

hemizygosity test is a direct way to identify specific genes responsible for phenotypic differences between strains with genetic variation (Stern, 2014). The F1 progenies of CB4856 x LJ1322 (*nta-1* deletion mutant of LJ1318 NIL) and LJ1321 (*nta-1* deletion mutant of CB4856) x LJ1318 NIL have the same heterogeneous genome for the QTL except for *nta-1*. They differ only at the *nta-1*, with CB4856 x LJ1322 carrying $nta\text{-}1^{CB4856}$ and LJ1321 x LJ1318 carrying $nta\text{-}1^{MY23}$. The F1 progenies of CB4856 x LJ1318 showed similar nictation fraction with CB4856 x LJ1322, suggesting CB4856 *nta-1* acts dominantly. Notably, the F1 progenies of LJ1321 x LJ1318 and CB4856 x LJ1322 showed significantly different nictation. This result indicates that the QTL effect of *nict-2* locus for nictation is dependent on variation in *nta-1*.

Sequence variation in the *nta-1* coding region affects predicted protein structures.

Natural variation in a gene can result in various changes, including changes in amino acids, splicing, and expression. It is probable that $nta\text{-}1^{MY23}$ had loss-of-function variations since its deletion had no effect. The variations in *nta-1* between CB4856 and MY23 were examined using the whole-genome sequencing data acquired from CeNDR. The *nta-1* locus had many natural variants, from its upstream promoter regions to the 3' UTR (Figure 12). Compared to CB4856 and

N2, MY23 had a short structural variation around the first intron and many SNVs in the coding regions. Sequencing of the PCR templates identified that the structural variation extended the first intron from 138 bp to 249 bp. Most SNVs in the coding regions were synonymous, and two were missense. 3,742,719 T>A induced Y216F and 3743476 A>G induced V55I amino acid changes in MY23. Neither change appeared to be in conserved domains, but the protein structures predicted using Colabfold suggested that the two structures might slightly differ around the C-terminal tail (Figure 14) (Mirdita et al., 2022).

Sequence variation in the *nta-1* promoter region induces variation in *nta-1* expression.

Next, I noted sequence differences in the promoter region. The expression patterns driven by the 0.85kb and 2.2kb upstream regions were checked (Figure 15). There were no differences in expression patterns dependent on their length, indicating that short promoters were sufficient. The promoters of both strains drove expression in a variety of tissues, including pharyngeal gland cells, pharyngeal muscles, body wall muscles, HMC (head mesodermal cells), and intestine. However, compared to the CB4856 promoter (*nta-1p*^{CB4856}), the expression pattern of the MY23 promoter (*nta-1p*^{MY23}) was not visible at all in head muscles and several cells around the middle of the pharyngeal isthmus and was relatively weak

in HMC and intestine (Figure 15, 16). This difference was also observed in L4, except that the expression in the head muscles was not well observed (Figure 17).

Cells around the pharyngeal isthmus did not overlay with the expression driven by the pan-neuronal promoter (*egl-3p::mCherry*), indicating that they were not neuronal cells (Figure 18). Aside from the neurons, the GLR cells were the most likely candidates based on their location and shape (Altun & Hall, 2009). The GLR cell has six cell bodies symmetrically located around the pharynx, just behind the nerve ring, and extends a sheet-like process anteriorly into the inner region of the nerve ring (Oikonomou & Shaham, 2011; White et al., 1986). The expression pattern of the fluorescence protein driven by GLR promoter (*nep-2p*) overlapped precisely with the expression driven by *nta-1p*^{CB4856} at the target cells, indicating that GLR cells are the site of differential expression by promoter variation (Figure 16). In addition, the expression of *nta-1* in the GLR cells was also identified in the N2-based single-cell transcriptome resource, CeNGEN (Table 2) (Taylor et al., 2021). Furthermore, according to the previously reported RNA-seq data (Zhang et al., 2022), *nta-1* expression was significantly lower in the wild strains carrying *nta-1p*^{MY23} (Figure 19). The lower expression level of the MY23 promoter group seemed consistent with the changes in expression patterns. These results show variants in *nta-1* promoter controls expression in various tissues, including GLR cells, head muscles, HMC, and intestine.

Sequence variation in the *nta-1* promoter controls nictation.

To determine whether sequence variation in the gene or the promoter are responsible for nictation, three *nta-1* transgenes with different combinations of promoter and gene were expressed in the *nta-1* mutant (Figure 20). In these transgenes, *sl2::gfp* was ligated to the end of *nta-1* to verify *nta-1* expression without interference with protein function. While expression of the transgene driven by *nta-1p*^{CB4856}, regardless of the gene sequence, reduced nictation of *nta-1* mutant similar to that of CB4856, expression driven by *nta-1p*^{MY23} had little effect (Figure 20). These results demonstrate that natural sequence variation in the *nta-1* promoter, rather than coding region, underlies the diversity of nictation.

***nta-1* expression in GLR cells regulates nictation.**

The above result that *nta-1* expression under *nta-1p*^{CB4856} specifically decreased nictation suggests that *nta-1* expression in additional tissues inhibited nictation. Among various tissues, I focused on the expression in GLR cells. GLRs are mesoderm-derived cells that cover the inner side of the nerve ring with their sheet structure and are sometimes referred to as glia (Singhvi & Shaham, 2019). GLRs form gap junctions with both head muscles and RME motor neurons (Altun & Hall, 2009), and the gap junctions between GLR and muscles were also found in

the electron miscopy images of a dauer (Figure 21) (Yim et al., 2023). In addition, the chemical synapse from IL2 neurons to GLR cells was found to be stronger in dauer stage (Figure 22) (Yim et al., 2023). As mentioned in introduction, IL2 neurons is essential for nictation (Lee et al., 2011). Given these, it was intriguing to speculate that *nta-1* expression in GLR cells may regulate nictation.

To confirm whether *nta-1* expression in the GLR cells underlies differences in nictation, the *nta-1* was expressed under previously described GLR cells promoters, *nep-2p* and *egl-6p* (Meng et al., 2016). Notably, *nta-1* expression induced by *nep-2p* and *egl-6p* appeared to reduce the nictation fraction of *nta-1* mutant to a level comparable to CB4856, implying that expression in GLR cell is sufficient to reduce nictation (Figure 23). At the same time, the more substantial effect of own promoter remained a possibility that expression in other tissues may also be involved. The data so far show that natural variation in the *nta-1* promoter causes differences in expression in GLR cells, resulting in the diversity of nictation.

***nta-1* is an ortholog of a human hydroxysteroid dehydrogenase.**

According to the descriptions available on the Wormbase, *nta-1* is predicted to have oxidoreductase activity and is an ortholog of *HSD17B14* (hydroxysteroid 17-beta dehydrogenase 14) of human (Davis et al., 2022).

Numerous proteins from various species were searched by BlastP of *nta-1* sequence in NCBI (Altschul et al., 1997) (Table 3), and various nematode orthologs were also present (Davis et al., 2022) (Table 4). *HSD17B14* is known to function in the metabolism of steroids at C17, for example, oxidizing estradiol to estrone (Lukacik et al., 2007). LET-767 of *C. elegans*, an ortholog of human *HSD17B12*, was proven to have a similar function with *HD17B12* when expressed in HEK-293 cells, converting estrone and androstenedione into estradiol and testosterone, respectively (Desnoyers et al., 2007). In the same way, NTA-1 may share a molecular function with *HSD17B14*, but this remains to be determined.

Geographic distribution of *nta-1* promoter haplotypes.

Currently, genetic information for 1525 strains grouped into 550 isotypes with different genetic features is available on CeNDR, enabling the population genetic analysis (Cook et al., 2017). To better understand the feature of the sequence variation in *nta-1* promoter, I examined the sequence of the *nta-1* promoter in 550 isotypes (Figure 24) and generated a tree based on their *nta-1* promoter sequences (Figure 25). Most isotypes shared sequences with CB4856 or MY23, with 409 categorized as CB4856 haplotype, 133 as MY23 haplotype, and the remaining 8 as other haplotypes. As when divided by the peak SNV, the group

with the MY23 haplotype showed higher nictation than the group with the CB4856 haplotype. (Figure 26).

Next, I examined the global distribution of *nta-1* haplotypes (Figure 27, Table 4). Globally, the frequency of the MY23 haplotype was 0.24 per isotype. However, there were significant regional characteristics. MY23 haplotype was concentrated in the Pacific region, including Hawaii and the American West (Figure 28). In Hawaii, the frequency of MY23 haplotype was 0.60 per isotype. In the American West, two haplotypes were present at half each. On the other hand, in the Europe and other regions, most isotypes have CB4856 haplotype, with MY23 haplotype present in only 0.02 and 0.05 per isotype. In other words, the diversity of *nta-1* haplotype has highly disappeared in other regions, but has been well conserved in the Pacific region, especially in Hawaii.

Diversity of *nta-1* promoter corresponds to ancient diversity in *C. elegans* populations.

It is noteworthy that Hawaiian isotypes show balanced haplotype distribution. The significance of diversity in *C. elegans* population of the Hawaii and the Pacific region has been investigated in several previous studies, which can explain *nta-1* haplotype distribution (Andersen et al., 2012; Crombie et al., 2019; Lee et al., 2021). A previous research showed that recent selective sweeps have

homogenized the genomes of *C. elegans* populations around, and the authors hypothesized that this may be related to the influence of human activity (Andersen et al., 2012). Another study found that the Hawaiian and non-Hawaiian *C. elegans* isotypes have distinct evolutionary histories, with Hawaiian isotypes having about three times higher genetic diversity than non-Hawaiian isotypes (Crombie et al., 2019). The authors suggested Hawaiian samples have been protected from the selective pressures assumed to be impacted by human activity in many part of the world and may display patterns of ancestral diversity (Crombie et al., 2019). Recent research has divided *C. elegans* populations into three groups (Global, Pacific, Hawaiian), with all European isotypes belonging to the Global group, while Hawaiian isotypes belonged to all three groups (Lee et al., 2021). The authors also suggested that the genomic regions in which this diversity is concentrated, the hyper-divergent regions, contain ancient diversity and have been maintained by long-term balancing selection (Lee et al., 2021). Taken together, the previous studies suggested *C. elegans* appeared to have originated in the Pacific regions, and the Hawaiian *C. elegans* population have maintained ancient diversity before the human-associated selective pressures.

A previous study that performed admixture analysis divided *C. elegans* populations into 11 groups, including 4 of Hawaiian population (Hawaiian Divergent, Volcano, Hawaiian Invaded, Hawaiian Low) (Crombie et al., 2019). Of

these, the Hawaiian Divergent (Hi-Div) and Volcanos group diverged early and avoided selective sweep or gene flow. I examined the distribution of *nta-1p* haplotypes in each Hawaiian population. Interestingly, the Hi-Div group, which was collected at higher elevations that have been less affected by human activities in Hawaii and had little admixture from other populations, already shared both haplotypes and showed balanced *nta-1p* haplotype frequency (Table 6a) (Crombie et al., 2019). Next, I checked the distribution of the Hawaiian isotypes in the latest genome-wide tree of 550 isotypes and re-grouped them in the tree. The balanced *nta-1p* haplotype distribution in the Hi-Div group that diverged early was identified again. (Figure 29, Table 6b). These results suggest that both *nta-1* promoter alleles have a very long history in the *C. elegans* population, and the diversity of *nta-1* promoter corresponds to ancient diversity.

nta-1 promoter alleles may have been maintained by balancing selection.

A closer look at the geographic distribution shows that the two old alleles coexist in a small area (Figure 28, Figure 30). The balanced distribution of nta-1 haplotypes in Hi-Div populations, a divergent group that has avoided sweep and gene flow, and the co-existence of two *nta-1* promoter haplotypes in a narrow region in Hawaii suggest the influence of balancing selection. I further validated this by calculating Tajima's D (Tajima, 1989), one of the ways to distinguish

selection processes for sequences including balancing selection (Greene, Brown, et al., 2016; Koenig et al., 2019). Tajima's D around the *nta-1* promoter showed substantially high values, supporting the ancestral balancing selection of two *nta-1* promoter haplotypes (Figure 31). Taken together, the diversity of *nta-1* promoter appears to be ancient genetic diversity that has been maintained by balancing selection.

Pleiotropy of *nta-1* on dispersal and reproduction may have shaped balancing selection of *nta-1*.

Higher nictation is expected to help increase fitness in nature as it can help the dispersal of dauer under harsh conditions. However, as described above, the *nta-1* haplotype seemed to have been maintained by balancing selection. Indeed, MY23 haplotype that increases nictation is not globally dominant, implying that CB4856 haplotype might have been favored in many regions with the higher human influence. Based on this, I hypothesized that variation in *nta-1* may affect not only nictation but also other phenotypes. At first, I examined the effect of *nta-1* on total fecundity, as exposure to estradiol increased reproduction rate of *C. elegans* (Tominaga et al., 2003). However, *nta-1* allele did not affect fecundity (Figure 32).

Next, I noted that steroid metabolism in *C. elegans* regulates both dauer entry and exit (Fielenbach & Antebi, 2008; Zhang & Sternberg, 2022). When dauers reach a favorable environment and resume development, the faster they start reproducing, the better they will be able to exploit the resources of the new environment and reproduce. To test this experimentally, I compared the initial brood size up to 48 hours after dauers were transferred to a plate with food. Notably, the deletion of *nta-1*^{CB4856} delayed post-dauer reproduction, resulting in a significantly lower brood size at 48 hours, while the deletion of *nta-1*^{MY23} had no effect (Figure 33a). This difference was not seen when comparing brood size in favorable conditions bypassing dauer, implying a specific effect of *nta-1* on post-dauer development (Figure 33b). In other words, *nta-1* shows pleiotropy on nictation of dauer and the post-dauer reproduction speed, with the CB4856 *NTA-1* inhibiting nictation while accelerating post-dauer reproduction, which may confer a trade-off between disperal and reproduction. These phenotypes may have shaped balancing selection of two *nta-1* promoter alleles. Although this result cannot explain the exact mechanism of the phenotype, it suggests that differences in reproductive speed can emerge in favorable environments. This could be due to differences in the threshold for dauer recovery or developmental speed after recovery. Meanwhile, deletion in NIL resulted in a higher brood size than in

CB4856, suggesting that the other gene in the QTL except *nta-1* also contribute to this phenotype.

The duplication of *nta-1*

Right next to *nta-1* is the *F12E12.12* pseudogene, which seems to have been duplicated recently and has a very similar sequence to *nta-1*. When aligned with *nta-1,* there are two different frameshift deletions of 1 bp and 2 bp in CB4856 and N2, respectively, resulting in a premature stop codon. On the other hand, in the MY23 genome, the entire 280 amino acid sequences can be retained in the same reading frame as *nta-1* without frameshift. In the RNA-seq results of the wild isolates (Zhang et al., 2022), it was identified that opposed to *nta-1,* the CB4856 group shows higher *F12E12.12* expression than the MY23 group (Figure 34). Although its expression is far low compared to *nta-1,* this data suggests a possibility that *F12E12.12* might have some function in the ALT group.

Discussion

Nictation is a conserved behavior in many nematode species that can facilitate interspecific interactions. In this study, I investigated the genetic basis of diversity in nictation among *C. elegans* wild isolates. I demonstrated that natural variation in promoters of *nta-1*, a putative oxidoreductase, underlies the diversity in nictation fraction by regulating its expression in various tissues. To the question of how behavioral diversity can emerge from genetic variation, this study provides a case that natural variation in the promoter of an enzyme for steroid metabolism can generate individual differences in the programmed dispersal behavior. In addition, population genetic analysis of *nta-1* showed that balancing selection, probably shaped by the pleiotropy of *nta-1*, has maintained the ancient diversity of *nta-1*. Taken together, this study will provide a research model for an integrated understanding of behavioral diversity.

Environmental contexts may have affected selection process of *nta-1* alleles

nta-1 affected both nictation and post-dauer reproduction speed. These two phenotypes may have dynamically affected the fitness of *C. elegans*, resulting in balancing selection or selection for the *nta-1* haplotype favored in each environmental context. In global regions other than Hawaii and the Pacific regions, most strains have the CB4856 *nta-1* promoter haplotype that accelerates post-dauer reproduction. This could be due to the human-associated influence, such as human-

assisted dispersal which can reduce the importance of dispersal by nictation and increase the advantage of faster post-dauer reproduction. In Hawaii, which has been less affected by human activity, dispersal through interaction with natural carriers may have been more significant than in other regions, shaping balancing selection for two ancient *nta-1* alleles. According to the strain information on CeNDR, Hawaiian *C. elegans* populations were collected mainly in the forest landscape (Cook et al., 2017), supporting this hypothesis.

Involvement of mesodermal systems in nictation

Nictation is a behavior in which the worm raises its head and body in three dimensions, distinct from the typical movement of other stages. It is anticipated that characteristic muscle coordination, such as synchronized muscle contractions, will be required, although it is unknown how the neuromuscular activity is regulated during nictation. It is remarkable that *nta-1* expression is different in mesodermal systems such as GLR and HMC, which are predicted to be involved in coordinating head and neck movements. Adding the recent finding that the synaptic connection from IL2 to GLR is stronger in dauer (Yim et al., 2023), I suggest a hypothesis that GLR cells act downstream of IL2 sensory neurons, regulating coordination of muscle activities involved in nictation.

Based on its connection with RME motor neurons and anterior muscles by gap junction and its location, GLR cells have been assumed to be involved in the coordination of neuromuscular junction activity and fine motor movements in the anterior body (Oikonomou & Shaham, 2011; Singhvi & Shaham, 2019; Stout Jr et al., 2014). Also, they have been presumed to be scaffolding cells guiding muscle arms (Altun et al., 2009). Head muscles are known to be the only muscle that have connections with glia including GLR, suggesting fine movement in head may require connections with the glia (Oikonomou & Shaham, 2011). In addition, HMC forms extensive gap junctions with both dorsal and ventral muscle arms and has been postulated to be involved in the synchronized contraction of head and neck muscles (Altun et al., 2009). GLR and HMC express diverse innexins, implying that gap junctions may be necessary for their function (Altun et al., 2009). Although there is no direct evidence that connections in these systems are important for nictation, I discovered that expression of *nta-1* driven by GLR promoters significantly reduced nictation. The findings and phenomena above provide a basis for further study about their potential roles in regulating nictation. Whether the connection between IL2 and GLR or the gap junction between GLR and muscles is important for nictation will be interesting research topics on the cellular basis of nictation. For example, studies using the IL2-ablated or

cholinergic transmission defect mutants will help to test whether the IL2-GLR connection is important for nictation.

Molecular function of *nta-1*

Although its role in nictation is identified in this study, the molecular function of *nta-1* and how it regulates nictation remain unclear. Steroid hormones affect many biological processes, including reproduction, development, and metabolism. They can act by binding to steroid receptors and affect gene expression, or by non-genomic way, activating signaling cascades (Lösel & Wehling, 2003). The most extensively investigated steroid metabolism in *C. elegans* involves dafachronic acids and their receptor, *daf-12*, which play a role in dauer formation, fat metabolism, developmental timing, and longevity (Antebi, 2015). Since I have confirmed that *nta-1* is involved in the reproduction rate after dauer, it would be helpful to understand its function by determining whether it is involved in the dauer recovery process or subsequent developmental processes.

Considering the existence of a functional estrogenic receptor *nhr-14* (Mimoto et al., 2007) and functional orthologs of *HSD17B12*, *let-767* in *C. elegans*, NTA-1 may play a similar role to HSD17B14. Further studies, such as the analysis of lipids from the mutant, and supplementation test of sterols in the mutant, will help to elucidate its molecular function. In addition, since *C. elegans* is dependent

on external cholesterol, it would be helpful to test the phenotypic differences with the mutant in cholesterol-deficient media to identify the function of NTA-1 in steroid metabolism.

It has been suggested that steroid hormones can regulate the expression of gap junction proteins (Michael Hendrix et al., 1995; Risek et al., 1990; Yu et al., 1994). Also, testosterone propionate and estradiol propionate can influence gap junction communication of Sertoli and cardiac cell of rats by direct membrane-steroid communication, disturbing properties of the cell membrane (Hervé et al., 1996; Pluciennik et al., 1996). Given the finding that *nta-1* is expressed in mesodermal systems where gap junctions are predicted to be important for their function, I suggest a hypothesis that *nta-1* expression may affect gap junction communication in these systems.

Significance of the variations in *nict-2* and around locus.

In this study, I have newly discovered that natural variation in *nta-1* regulates both nictation and post-dauer reproduction. Interestingly, the significance of natural variation in *nict-2* and around has been independently reported in other studies. As shown in Table 1, the *nict-2* QTL contains many variants that may affect the function of genes other than *nta-1*. A previous study that examined QTL for dauer development using RIL discovered a QTL on chromosome II (3.66 – 3.87

Mb) that covers *nict-2* QTL (3.71 – 3.80 Mb) (Harvey et al., 2008). DR1350 strain, collected in California and carrying the MY23 *nta-1*, was used to make NIL crossed with N2. The genome of DR1350 reduced dauer formation, implying the ALT genome of or around *nict-2* contributes to reduced dauer formation. Another study reported increased *nta-1* (*F12E12.11*) expression under dauer-inducing conditions (Harvey et al., 2009). In addition, another study has shown that hyper-divergent regions contain many environmentally responsive genes, with the locus containing many GPCRs (II:3,667,179 - 3,701,405), right next to the *nta-1* being a prime example. (Lee et al., 2021). Given these, *nict-2* and around might be an example of a locus with enriched diversity regulating many dauer-related phenotypes.

Figures and Tables

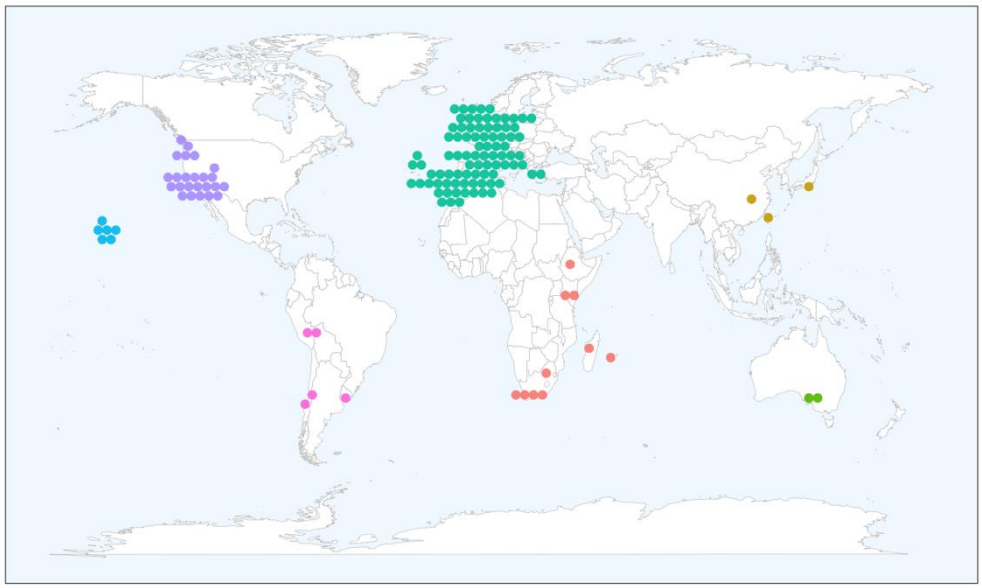

Figure 1. Worldwide distribution of *C. elegans* wild isolates.

Global distribution of the 135 of 139 wild isolates used in this study. Each dot indicates a strain. Dots are colored by continents except for Hawaii (blue). (purple, North America; pink, South America; green, Europe; orange, Afreeca; yellow, Asia; light green, Australia)

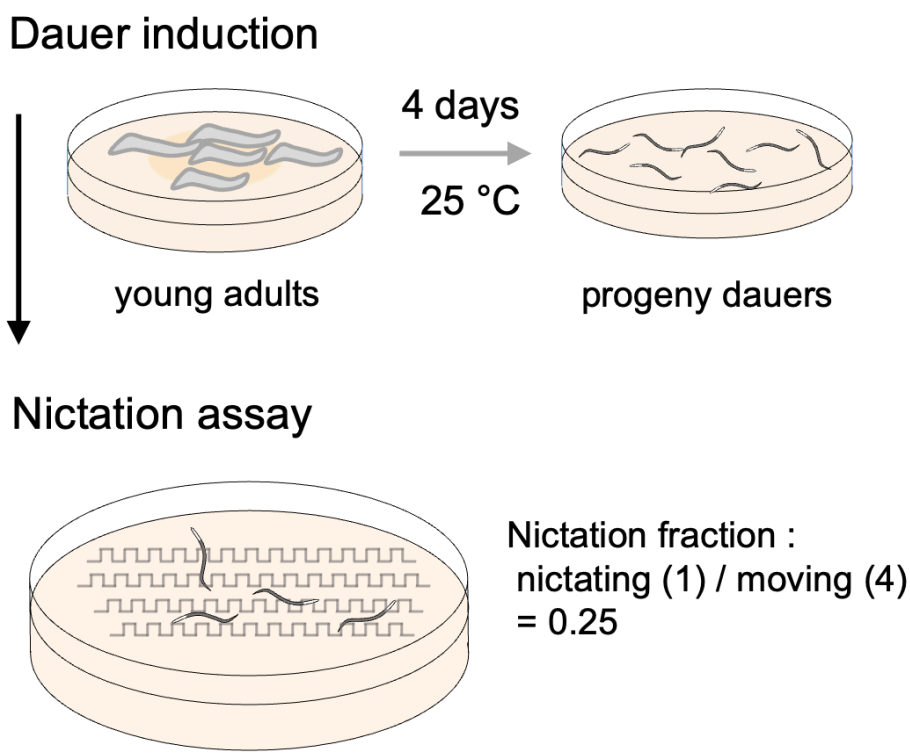

Figure 2. Schematic of nictation assay

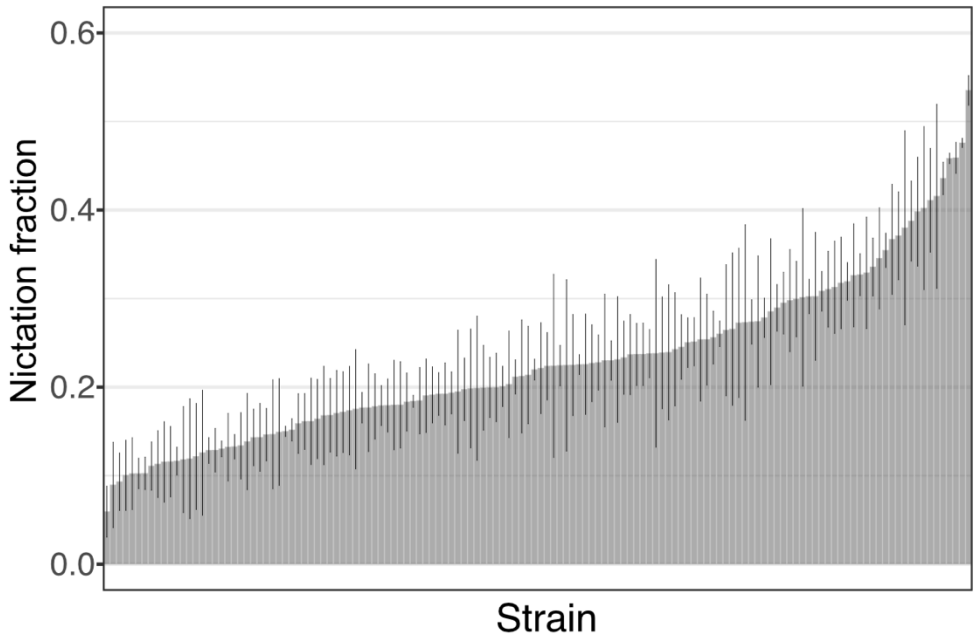

Figure 3. Nictation fraction is different among wild isolates.

Nictation fraction of 139 wild isolates. All strains have three replicates. Error bars show standard deviation.

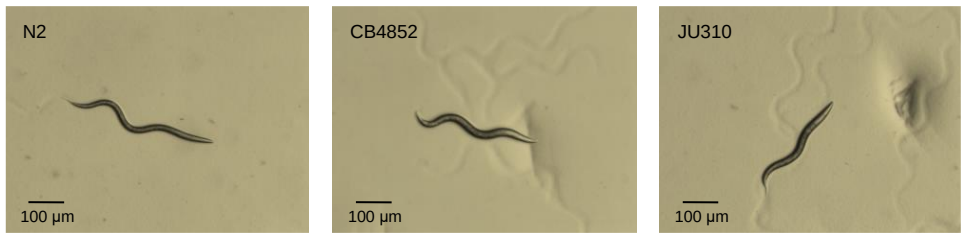

Figure 4. CB4852 and JU310 strains display different dauer morphology.

Pictures of the dauer of N2 (left), CB4852 (middle), JU310(right).

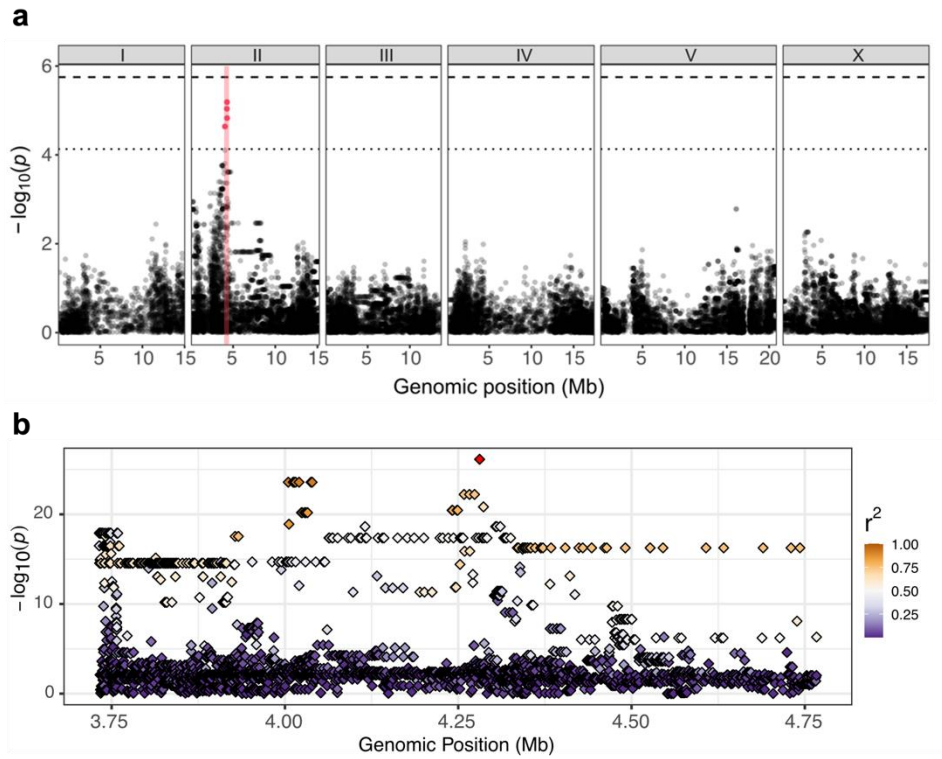

Figure 5. Genome-wide association mapping revealed a QTL on chr II.

(a) Result of the GWA mapping of the nictation fraction. The upper horizontal dash line corresponds to the Bonferroni-corrected threshold using all markers, and the lower horizontal dash line corresponds to the EIGEN threshold, which is the Bonferroni-corrected threshold correcting for the number of independent markers (genome-wide eigen-decomposition threshold). SNVs are colored red if they pass the second threshold. A vertical red line represents the QTL on chromosome II.

(b) Fine mapping of the QTL. Fine mapping was performed by evaluating the genotype-phenotype relationship for variants nearby the QTL identified from GWA mapping.

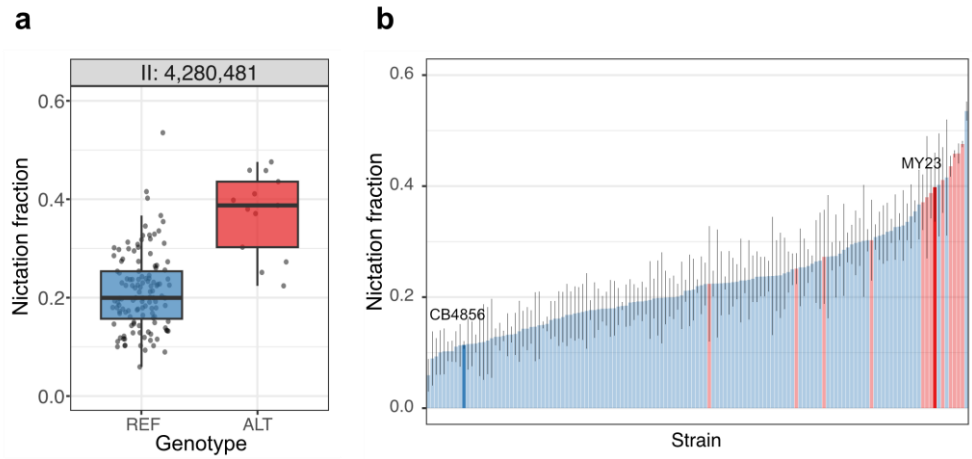

Figure 6. ALT group shows higher nictation fraction than REF group.

(a) Phenotype by genotype split for the QTL region. The entire strains are divided into two groups, with the peak SNV as a marker. REF indicates the reference group that have same SNV with N2, and ALT indicates the alternative group that have the other base.

(b) Colored version of Figure 3, based on the genotype of peak SNV. CB4856 and MY23, strains that used as representatives of REF and ALT group for the generation of NILs, are highlighted.

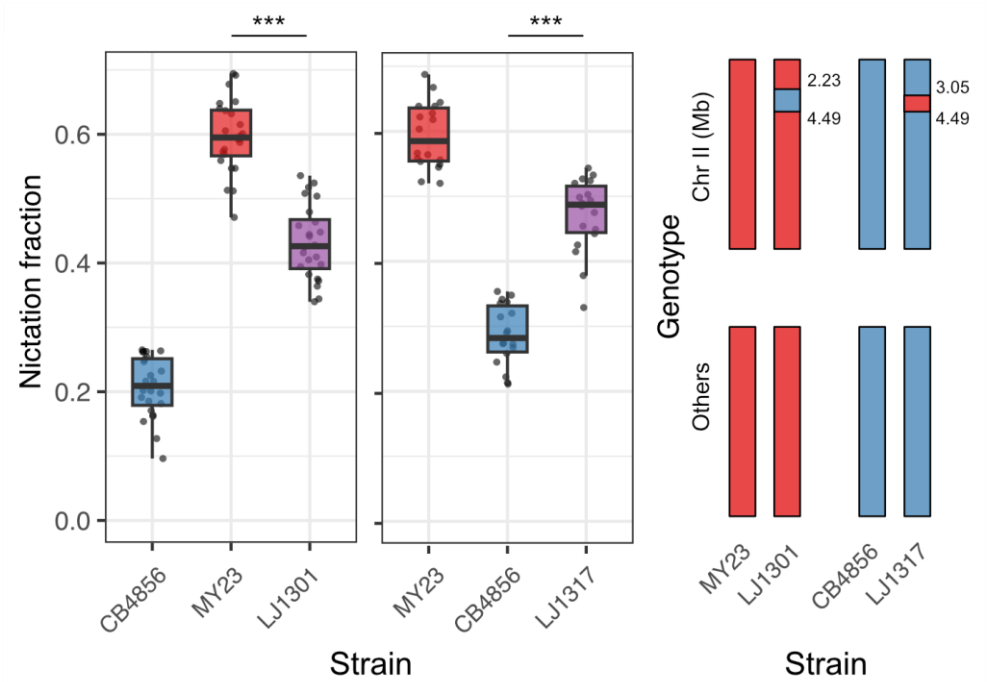

Figure 7. The QTL effect is identified in bi-directional NIL.

(left) Nictation fraction of two NILs, LJ1301(*nict-2*, CB4856>MY23) and LJ1317(*nict-2*, MY23>CB4856). Statistics: ***p<0.001. Pairwise t-test.

(right) Genotypes of two NILs, LJ1301 and LJ1317. Blue and red color represent the genome of CB4856 and MY23, respectively.

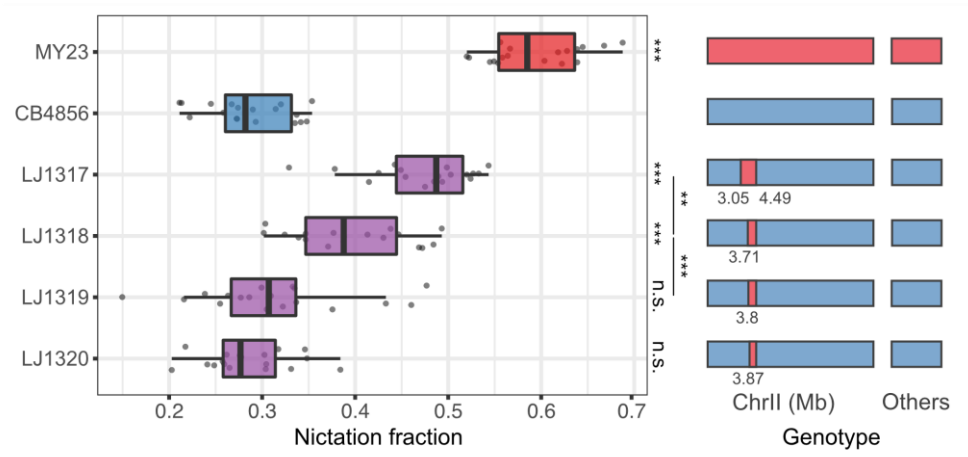

Figure 8. Fine mapping of *nict-2* QTL narrows its interval to 90 kb.

(a) Nictation fraction of NILs with CB4856 background (*nict-2*, MY23>CB4856) LJ1317, LJ1318, LJ1319, LJ1320. Statistics: n.s.: non-significant, ***p<0.001, ** p<0.05. Pairwise t-test.

(b) Genotypes of NILs. Blue represents the CB4856 genome, and red represents the MY23 genome.

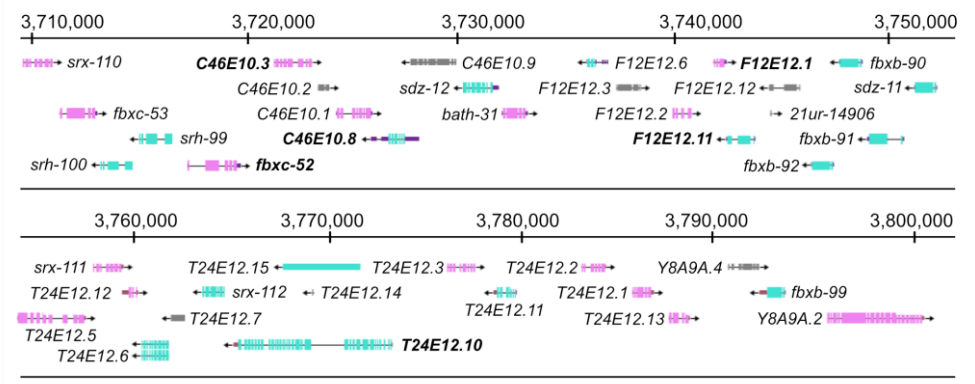

Figure 9. A 90 kb *nict-2* QTL.

A 90 kb *nict-2* QTL contains 31 protein-coding genes, 5 pseudogenes, 2 ncRNAs, and a 21U-RNA gene. Protein coding genes are colored pink or cyan based on their directions. Others are colored gray. The image obtained from JBrowse 2 of Wormbase was modified to generate this figure.

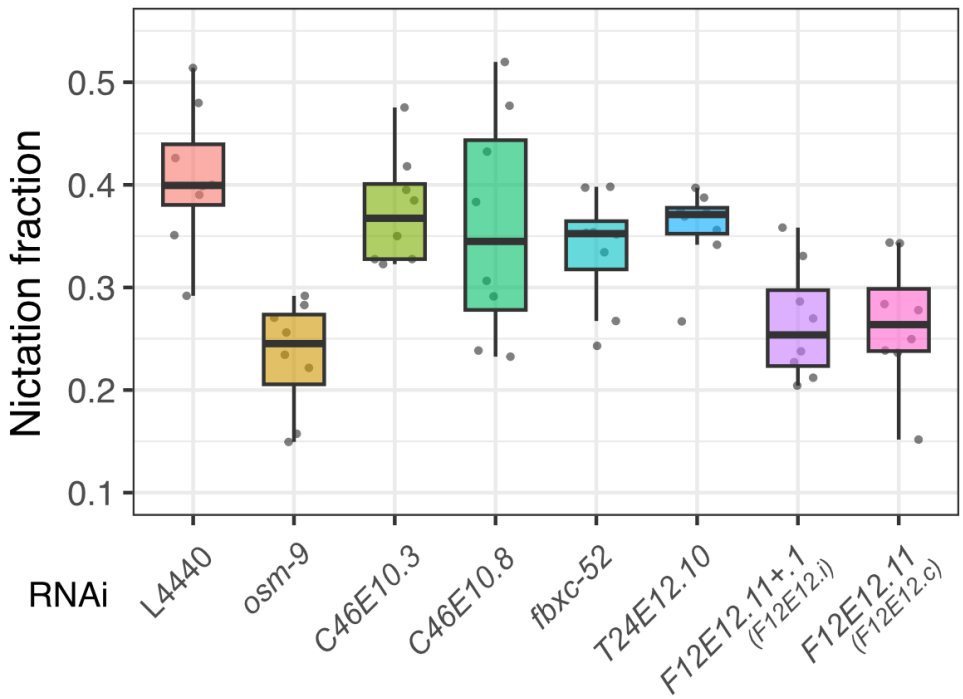

Figure 10. RNAi of *F12E12.11* affects nictation.

Nictation fraction of RNAi experiments for several candidate genes. *osm-9* was used as a control of RNAi experiments, as its null mutation is known to reduce nictation.

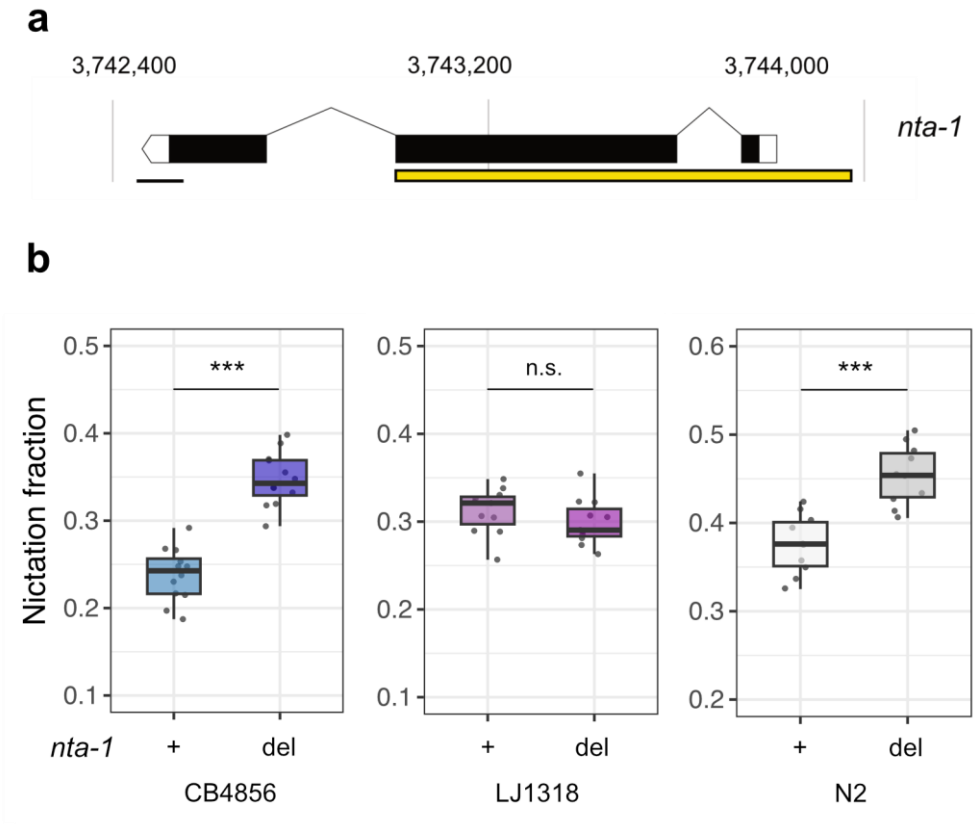

Figure 11. *nta-1(F12E12.11)* deletion in REF group increases nictation.

(a) Genetic structure of *nta-1*. The yellow box represents the CRISPR-Cas9 mediated deletion site.

(b) Nictation fraction of *nta-1* mutants of CB4856, LJ1318, N2. Statistics: n.s.: non-significant, ***: p<0.001. Pairwise t-test.

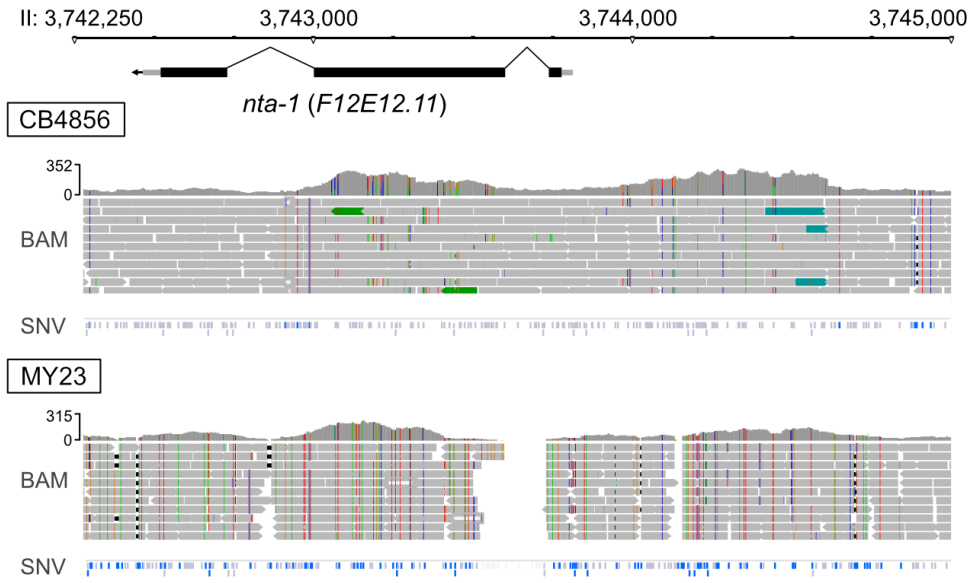

Figure 12. Variants in *nta-1*

BAM: Short-read alignments of CB4856 and MY23 to the N2 reference genome (WS245) around *nta-1*.

SNV: SNVs of each strain compared to the N2 reference genome are colored blue.

The image obtained from CeNDR was modified to generate this figure.

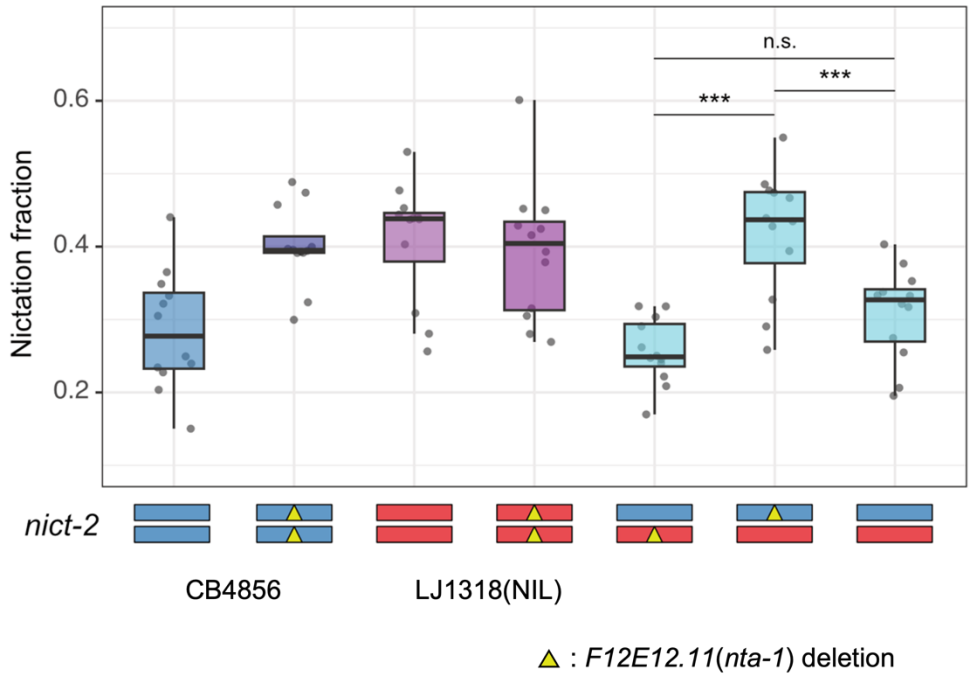

Figure 13. Variation in *nta-1* underlies *nict-2* QTL effect on nictation.

Nictation fraction of *nta-1* mutants and reciprocal F1 progenies. CB4856, LJ1318, and *nta-1* mutants of both strains were used. The bottom of the figure shows the genotypes of *nict-2*. All except *nict-2* shown in this picture are composed of the CB4856 genome. Blue and red box represent the *nict-2* genome of CB4856 and MY23, respectively. Yellow triangle represents *nta-1* deletion. Statistics: n.s.: non-significant, ***: $p<0.001$. Pairwise t-test.

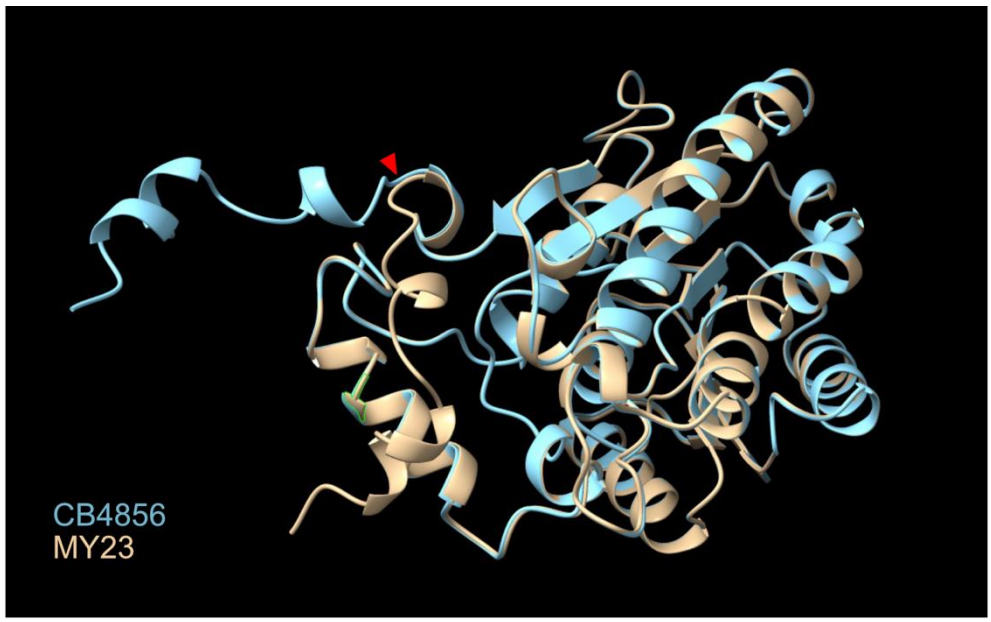

Figure 14. Predicted structures of NTA-1.

Only a top-scored structure of each sequence is displayed. Red triangle marks the point where two predicted structures diverge. Structures were predicted using ColabFold v1.5.2 and visualized using Chimera X.

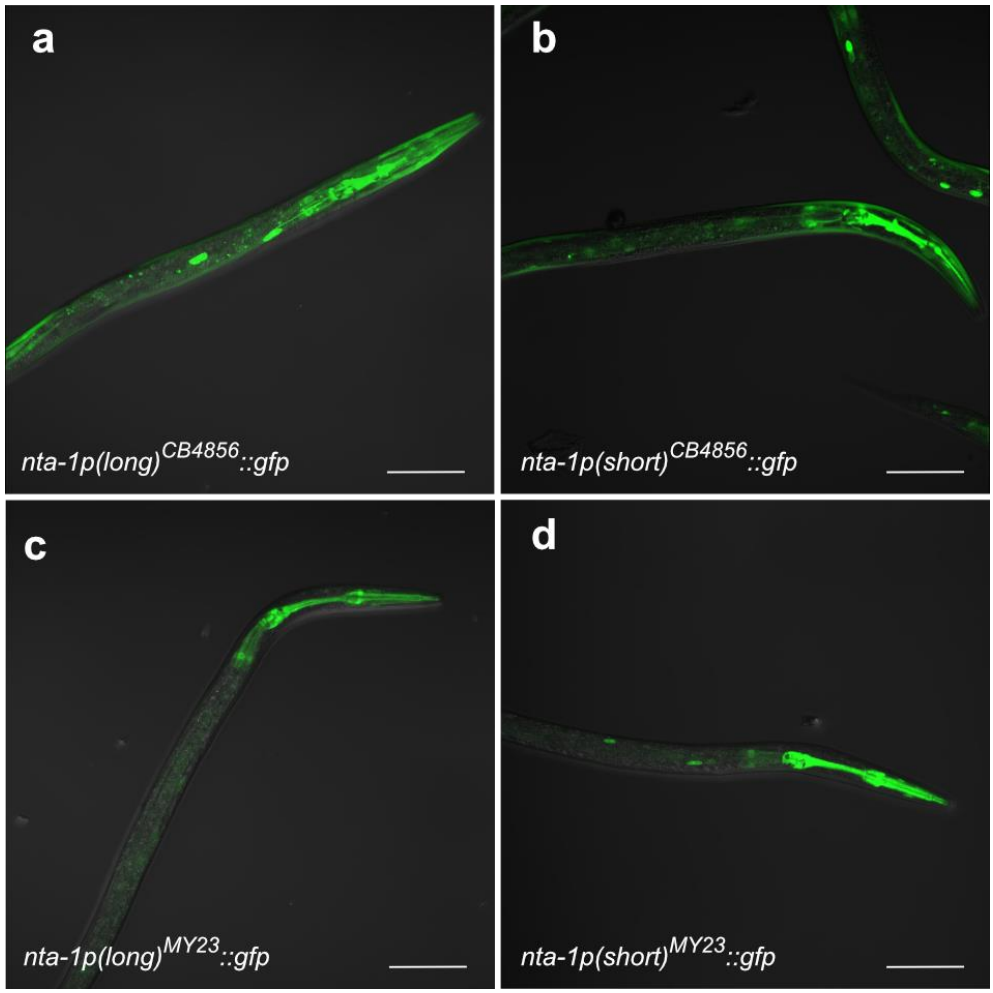

Figure 15. Different expression patterns driven by CB4856 and MY23 promoters.

(a,b) Expression patterns in N2 dauers driven by long(a) and short(b) *nta-1p*^{CB4856}.

(c,d) Expression patterns in N2 dauers driven by long(c) and short(d) *nta-1p*^{MY23}.

Scale bars: 50 μm.

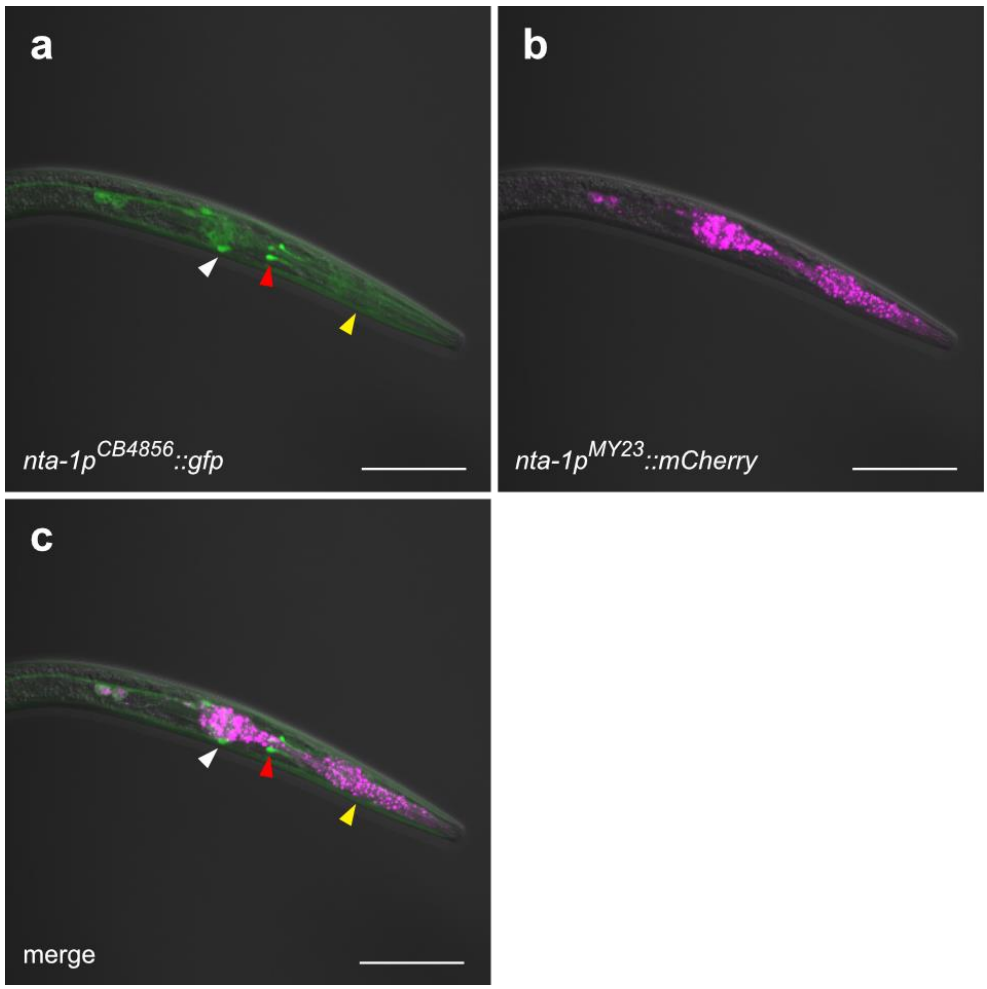

Figure 16. Expression variation in head muscles, head mesodermal cell, and other cells near the nerve ring.

(a-c) Maximum intensity projection images of expression patterns in N2 dauers driven by *nta-1p*^{CB4856}(a) and *nta-1p*^{MY23}(b). (c)merge. Yellow and white arrows show expression in head muscles and head mesodermal cells. Red arrows show expression in the middle of pharyngeal isthmus, near the nerve ring. Aggregation of mCherry occurred under dauer induction conditions. Scale bars: 50 μm.

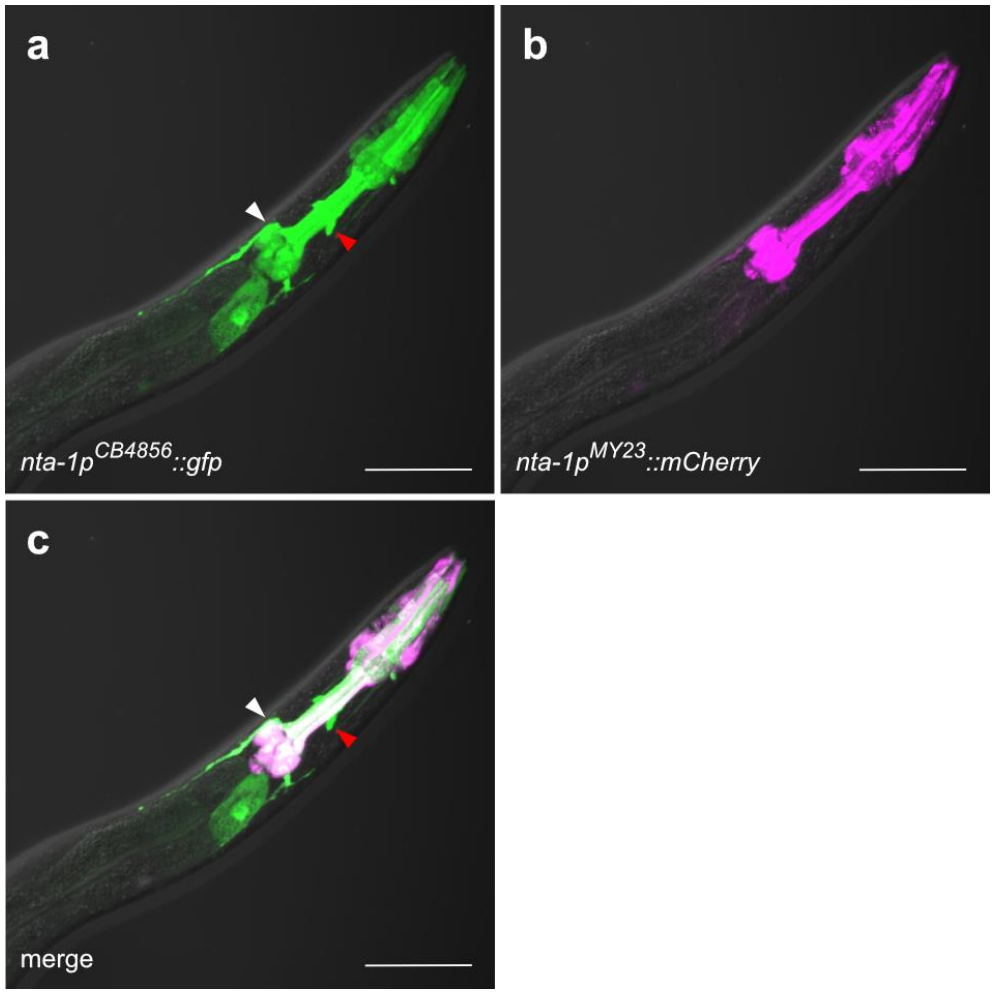

Figure 17. Expression patterns in L4 worms.

(a-c) Maximum intensity projection images of expression patterns in N2 L4 driven by *nta-1p*^{CB4856}(a) and *nta-1p*^{MY23}(b). (c)merge. White arrows show expression in head mesodermal cells. Red arrows show expression in the middle of pharyngeal isthmus. Scale bars: 50 μm.

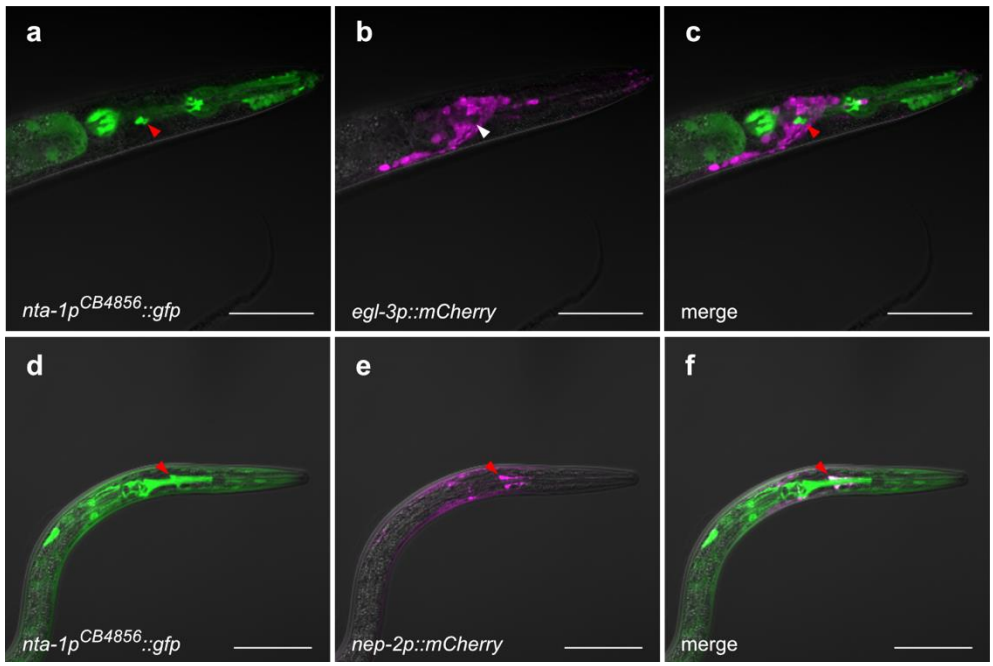

Figure 18. Expression variation in GLR cells.

(a-c) A section showing expression patterns driven by *nta-1p*^{CB4856}(a) and *egl-3p*(b) in an L4 worm. (c)merge.

(d-f) A section showing expression patterns driven by *nta-1p*^{CB4856}(a) and *nep-2p*(b) in a dauer. (c)merge.

A short *nta-1p* was used for *gfp* expression. For comparison to neuronal expression (*egl-3p::mCherry*), An L4 worm was used to avoid mCherry aggregation under dauer induction conditions. Red arrows show expression in GLR cells. White arrows show same location without expression. Scale bars: 50 μm.

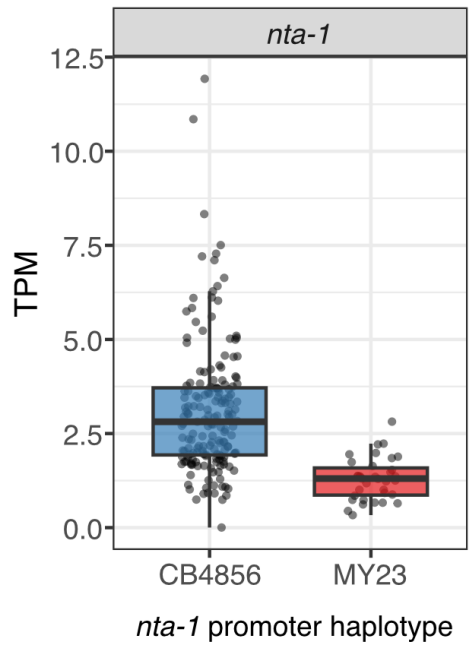

Figure 19. *nta-1* expression of wild isolates by *nta-1* promoter allele.

Phenotype by genotype split for *nta-1* TPM in the RNAseq data of wild isolates. Each dot represents a strain. RNAseq data were obtained from Zhang et al. (Zhang et al., 2022). *nta-1* promoter haplotype will be discussed later in Figure 24 and 25.

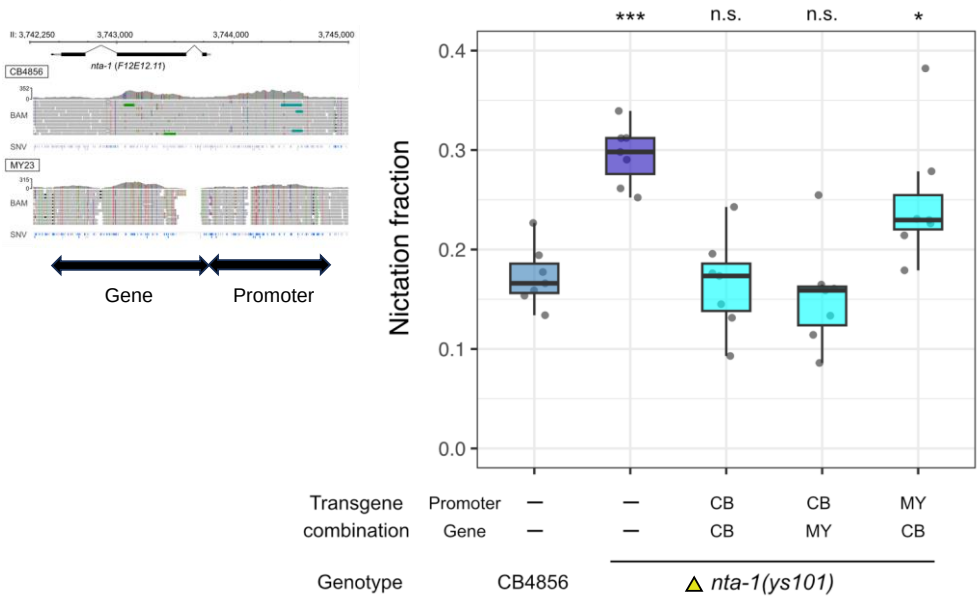

Figure 20. Variants in *nta-1* promoter regulates nictation.

Nictation fraction of worms expressing the *nta-1* transgene in different combinations. CB4856(first column), and LJ1321(second column) were tested with transgenic worms expressing *nta-1* of different promoter and gene combinations. CB and MY indicate sequence from CB4856 and MY23, respectively. *nta-1* transgenes were expressed in LJ1321. Statistics: n.s.: non-significant, ***: $p<0.001$, *: $p<0.05$. Pairwise t-test.

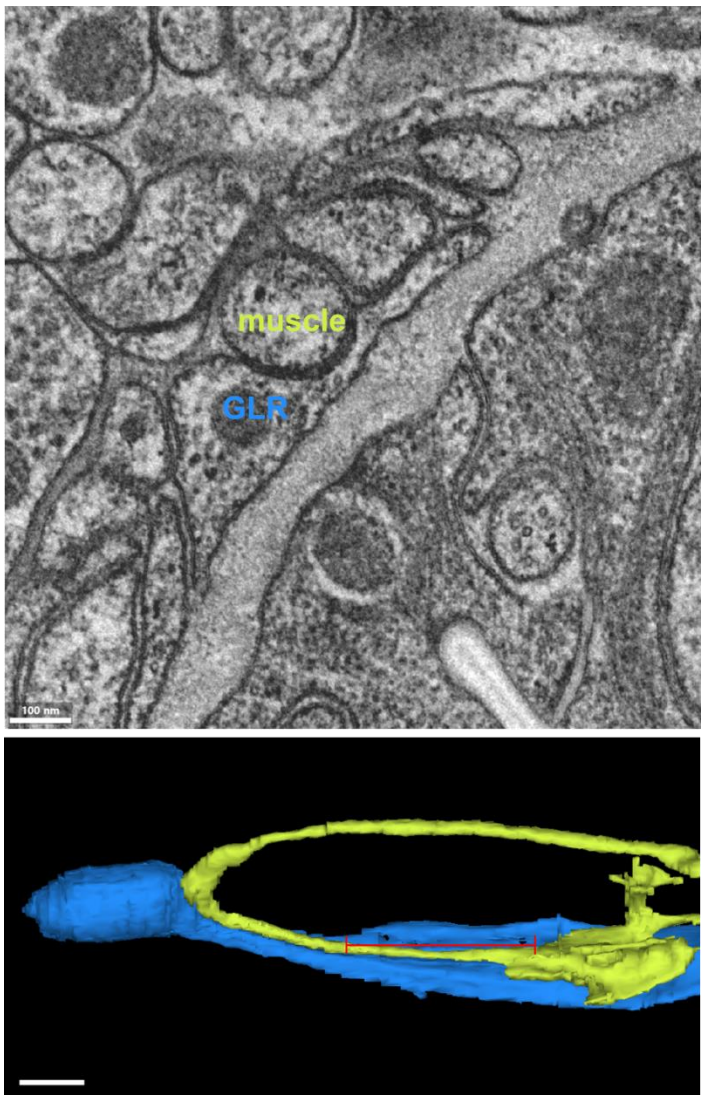

Figure 21. Gap junctions between GLR and muscle in dauer.

(top) An electron microscopy image of cross section of a dauer that shows gap junction between GLR and muscle. Scale bar: 100 nm.

(bottom) Reconstructed 3D structures of a GLR(GLRDR) and a muscle arm. Gap junctions are mostly found in the range marked in red. Scale bar: 1 μm.

In collaboration with Yim et al. (Yim et al., 2023).

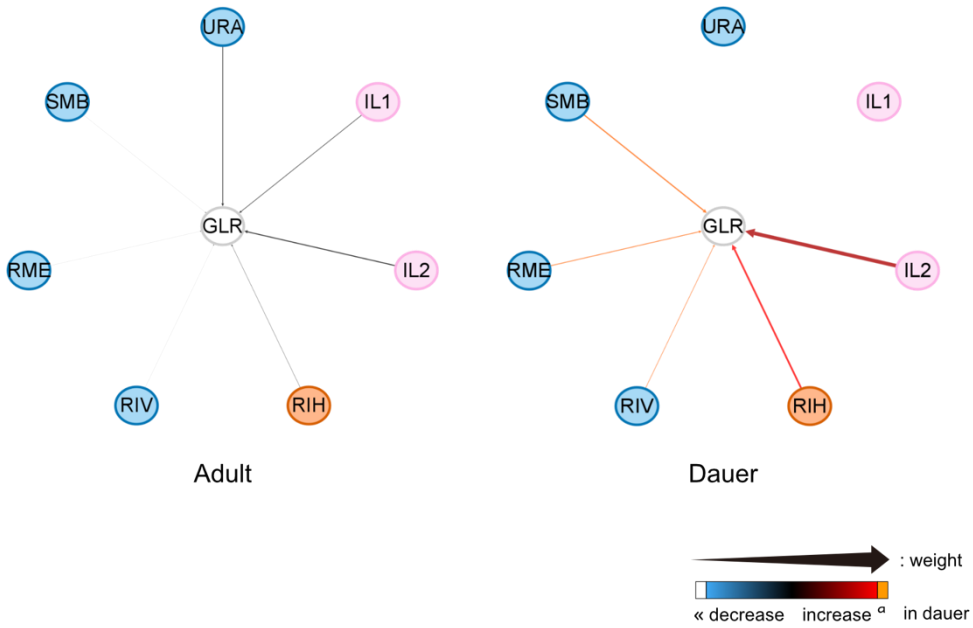

Figure 22. Chemical connection to GLR from IL2 is stronger in dauer than in adult.

Diagrams showing the chemical synapses of GLR cells in adult(left) and dauer(right). In collaboration with Yim et al. (Yim et al., 2023).

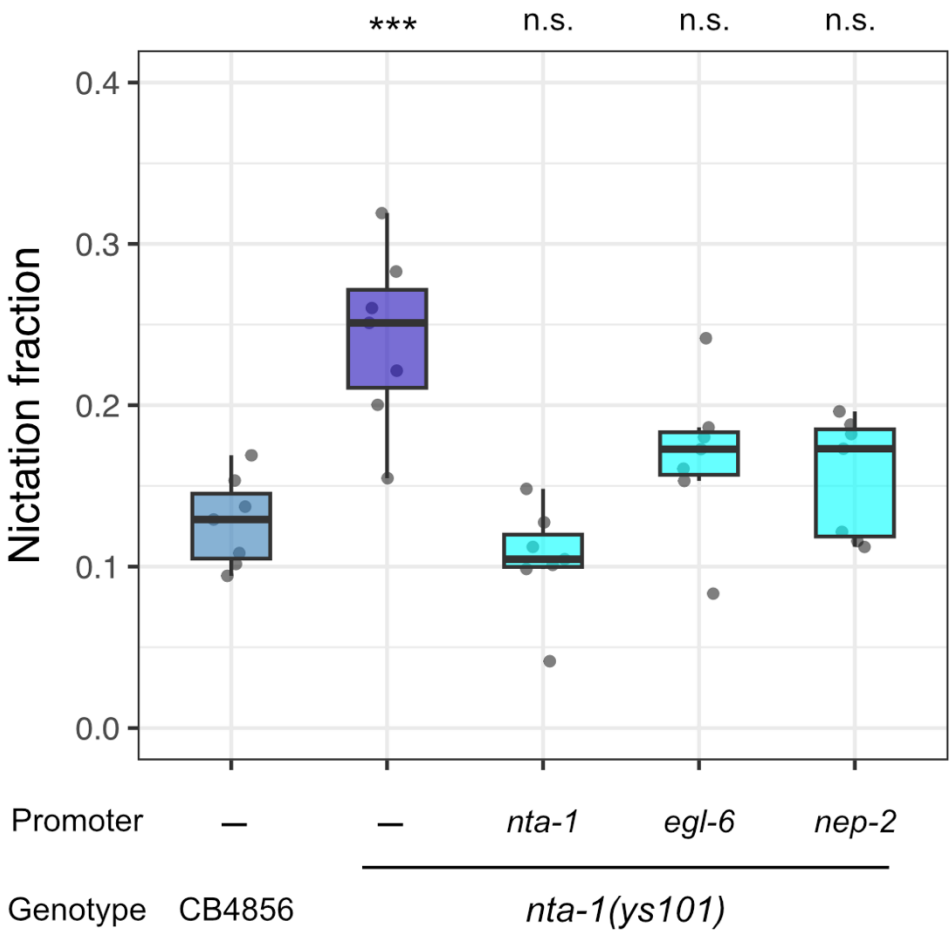

Figure 23. Expression of *nta-1* in GLR cells is sufficient to reduce the nictation of *nta-1* deletion mutant.

CB4856(first column), and LJ1321(second column) were tested with transgenic worms expressing *nta-1* using different promoter and gene combinations. Statistics: n.s.: non-significant, ***: p<0.001. Pairwise t-test.

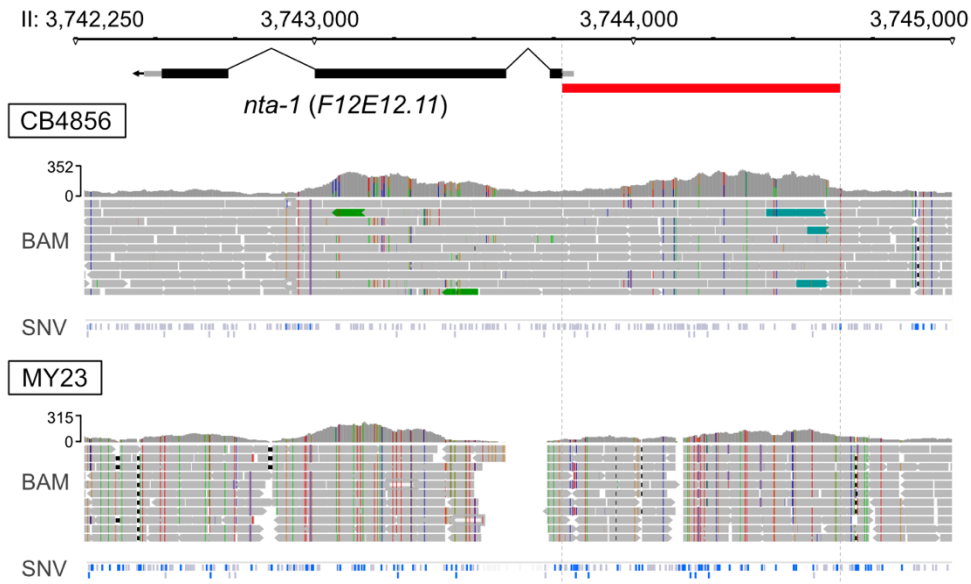

Figure 24. The genomic region for haplotype search.

Sequences at the *nta-1* promoter locus (II: 3,744,633-3,743,777 in N2 reference genome), shown in red box, were analyzed. The image obtained from CeNDR was modified to generate this figure.

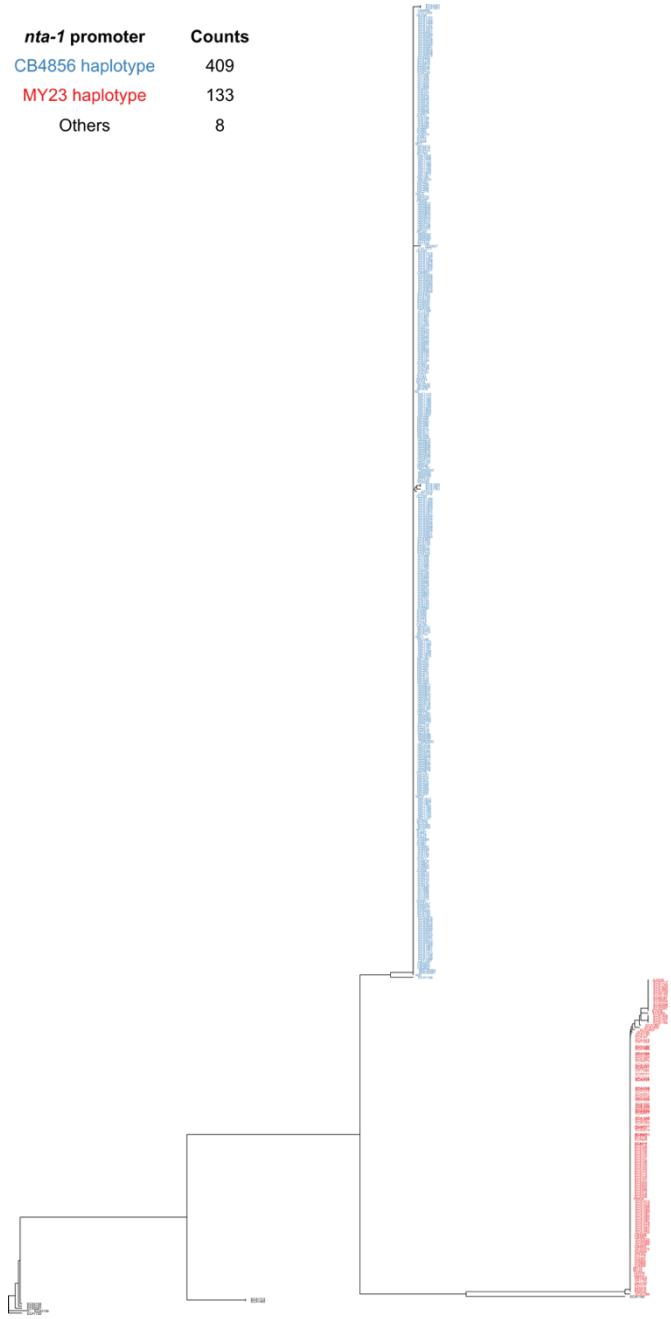

Figure 25. A tree of the *nta-1* promoter of 550 *C. elegans* isotypes.

Isotypes that have *nta-1* promoter haplotype of CB4856 and MY23 are shown in blue and red, respectively. In collaboration with D. Lee.

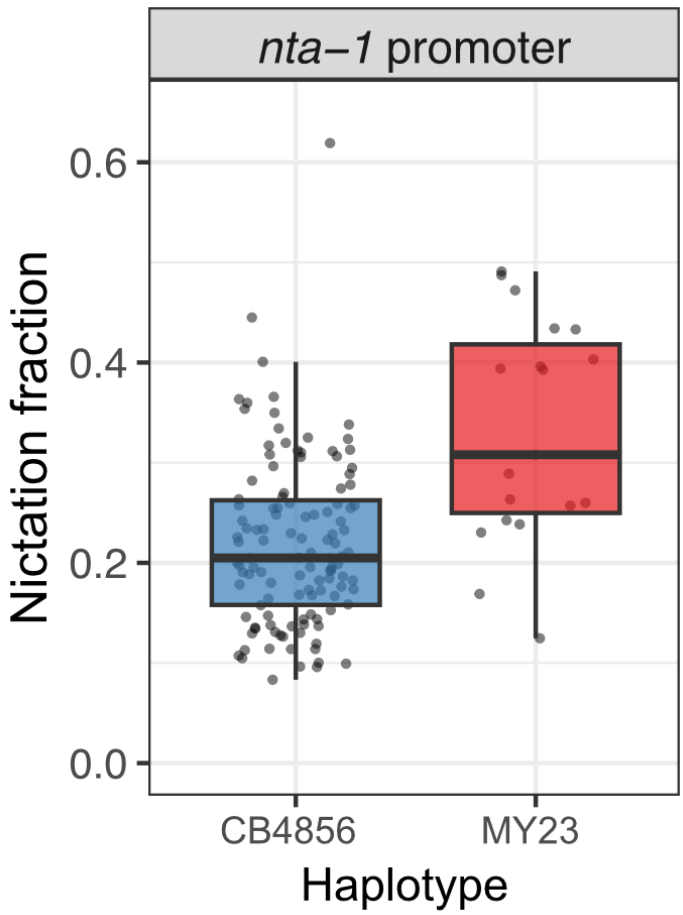

Figure 26. Nictation fraction of wild isolates by *nta-1* promoter haplotype.

Phenotype by genotype split for nictation fraction of wild isolates. Strains are split based on the haplotype of *nta-1* promoter.

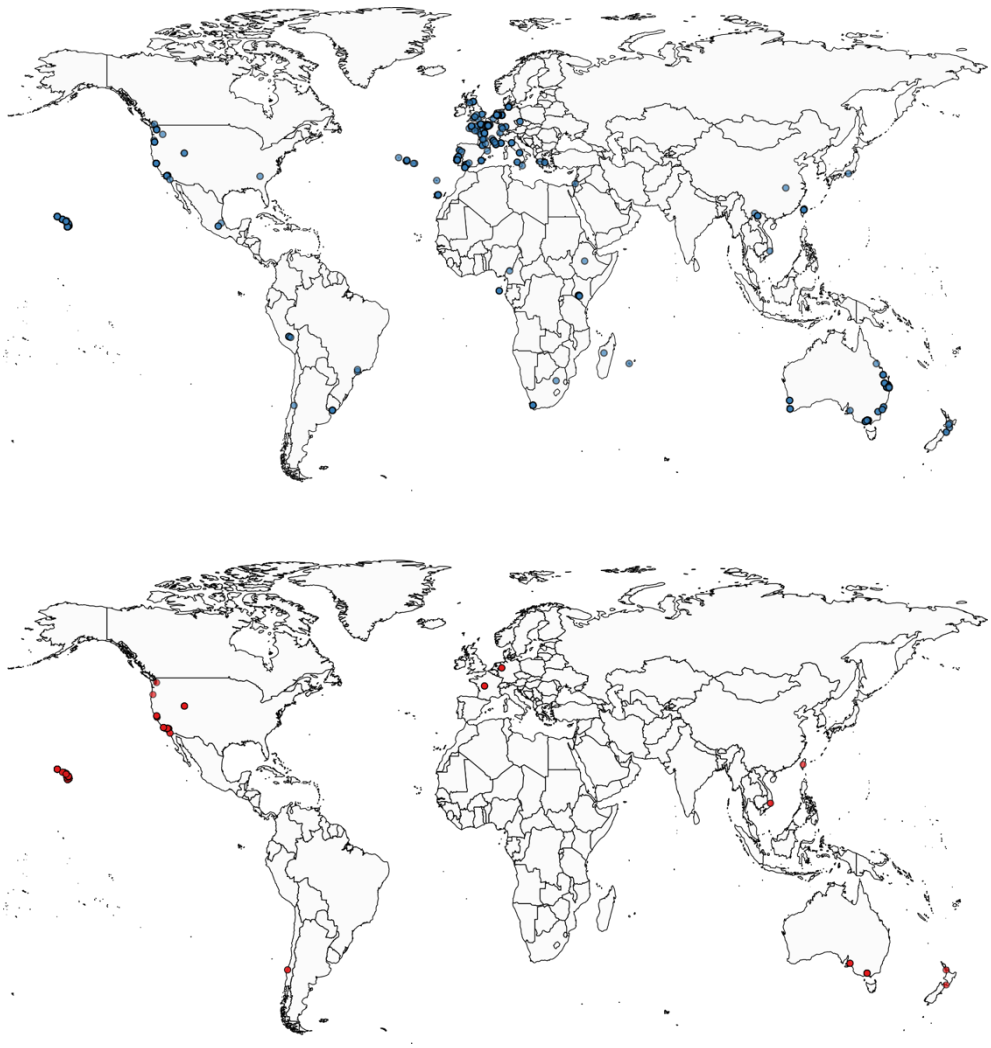

Figure 27. Global geographic distribution of *nta-1* promoter haplotypes.

(top) Global distribution of the CB4856 *nta-1* haplotype (blue).

(bottom) Global distribution of the MY23 *nta-1* haplotype (red).

Each dot represents a strain.

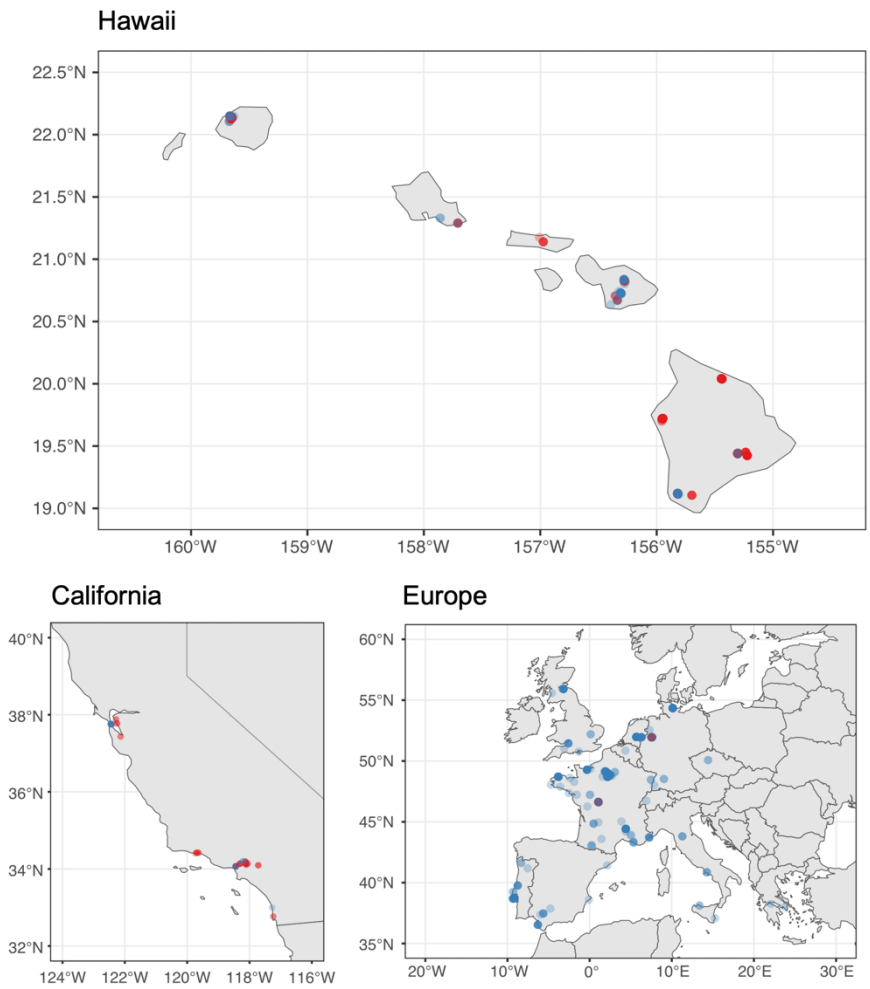

Figure 28. MY23 *nta-1* promoter haplotype is concentrated in Hawaii and Pacific regions.

Distribution of the CB4856 (blue) and MY23 (red) *nta-1* haplotypes in Hawaii, California, and Europe.

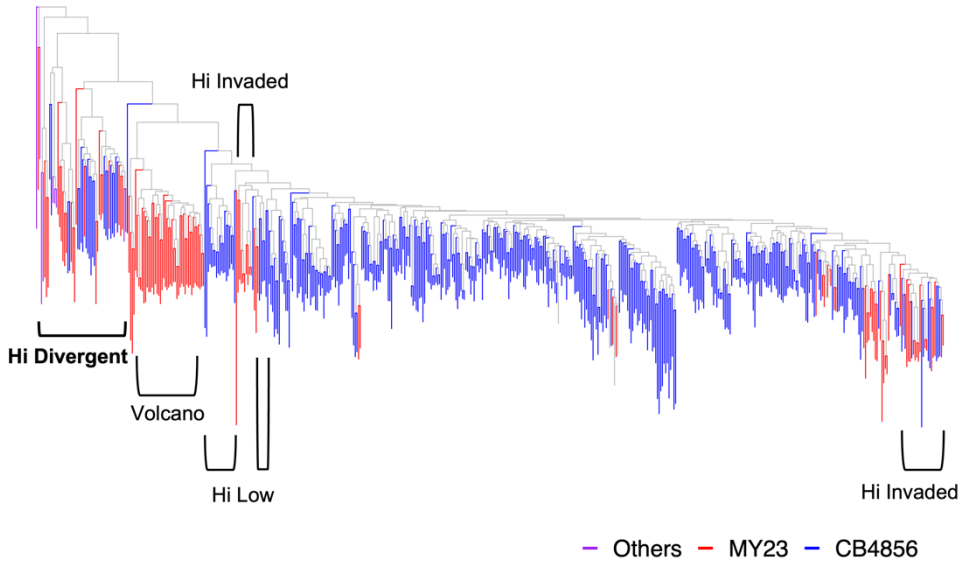

Figure 29. Distribution of *nta-1* promoter haplotype in a genome-wide tree of 550 *C. elegans* isotypes.

Isotypes that have *nta-1* promoter haplotype of CB4856 and MY23 are shown in blue and red, respectively. Four Hawaiian groups were reconstructed based on the range of the distribution of isotypes in previously divided groups in Crombie et al. (Crombie et al., 2019). In collaboration with D. Lee.

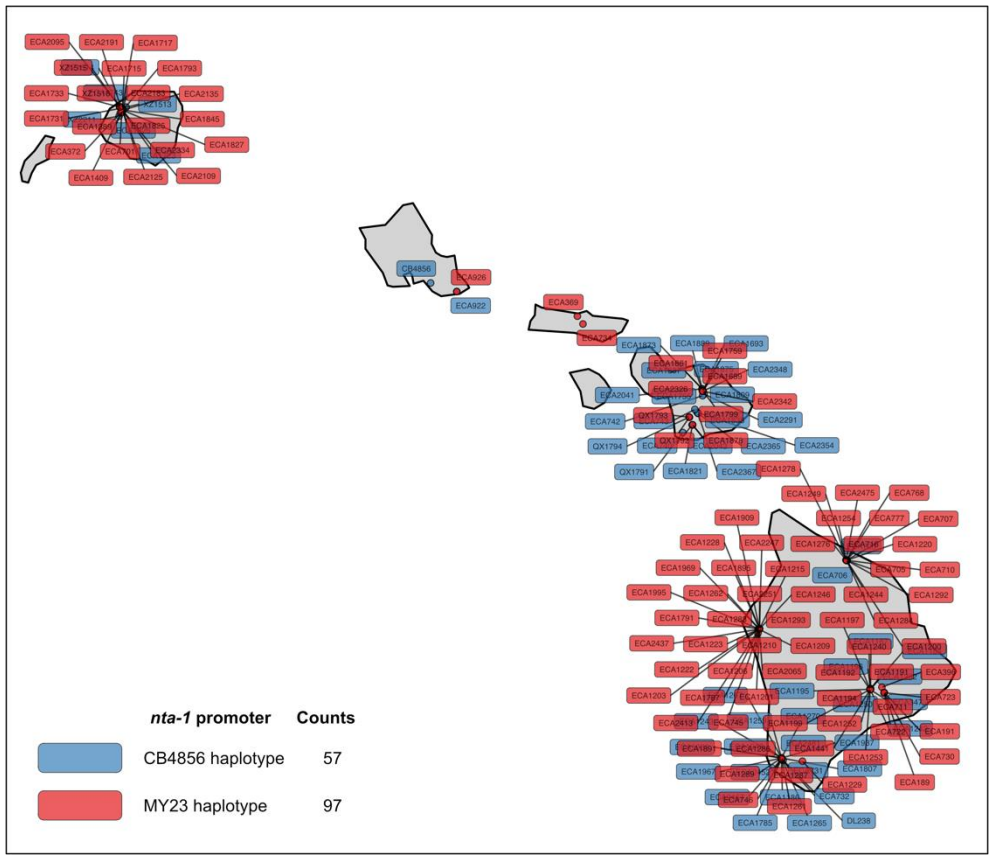

Figure 30. Two *nta-1* promoter haplotypes coexist in a small area in Hawaii.

Geographic distribution of the *nta-1* promoter haplotype in Hawaii.

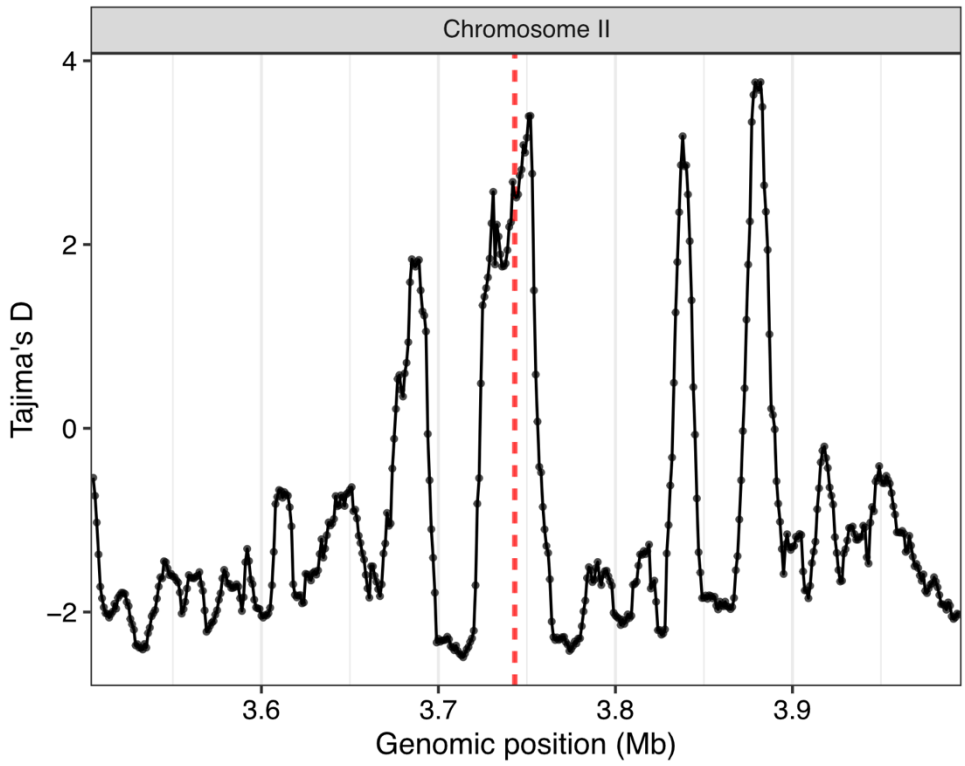

Figure 31. Tajima's D at the *nta-1* locus supports balancing selection.

Tajima's D statistics across the *nta-1* locus. Dashed red line indicates the location of *nta-1*. In collaboration with D. Lee.

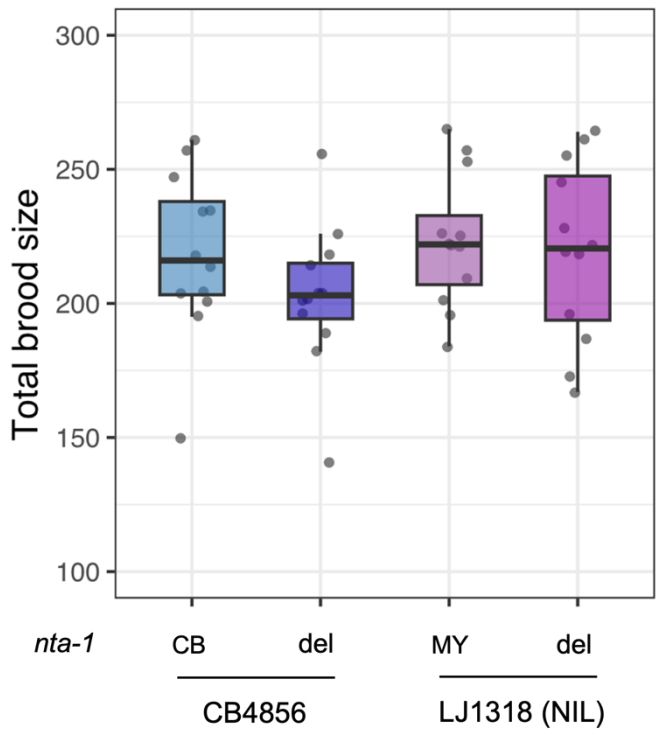

Figure 32. *nta-1* does not affect total brood size.

Total brood sizes of CB4856, LJ1318 (*nict-2*, MY23>CB4856), and their respective *nta-1* deletion mutants.

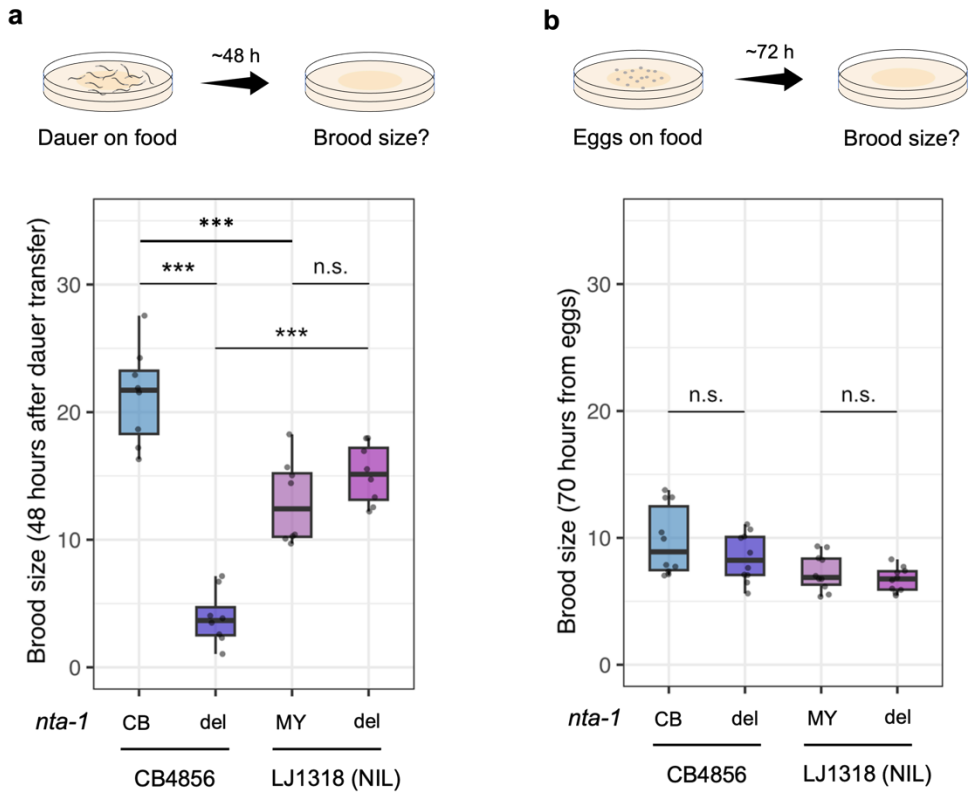

Figure 33. *nta-1* influences post-dauer reproduction speed.

(a) Brood sizes of CB4856, LJ1318 (*nict-2*, MY23>CB4856), and their respective *nta-1* deletion mutants at 48 hours after dauer transfer.

(b) Brood sizes of CB4856, LJ1318 (*nict-2*, MY23>CB4856), and their respective *nta-1* deletion mutants at 70 hours from eggs.

Statistics: n.s.: non-significant, ***: p<0.001. Pairwise t-test.

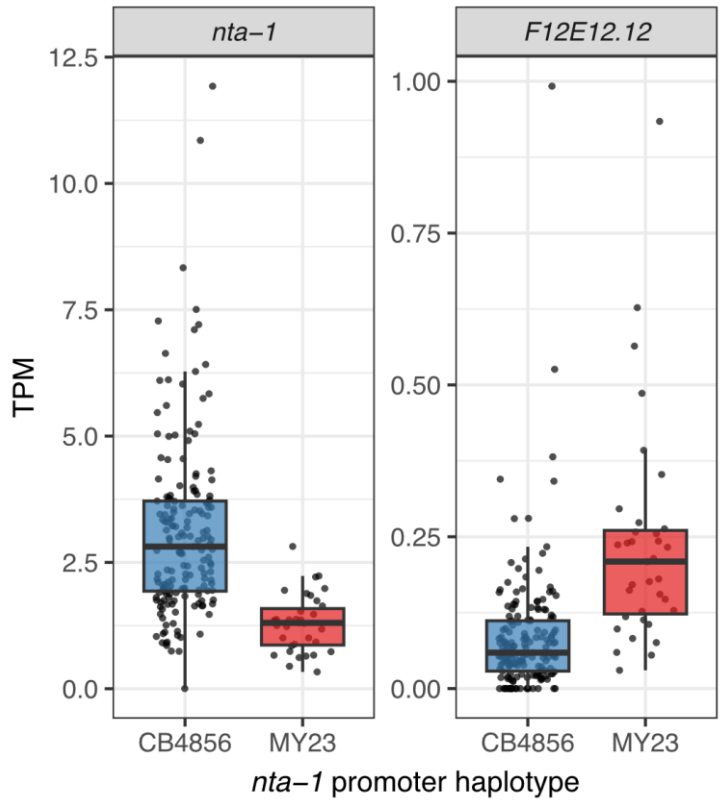

Figure 34. Expression of *nta-1* and *F12E12.12* of wild isolates by *nta-1* promoter haplotype.

Phenotype by genotype split for *nta-1* and *F12E12.12* TPM from the RNAseq data of wild isolates obtained from Zhang et al. (Zhang et al., 2022). Each dot represents a strain.

Table 1. Variants between CB4856 and MY23 in *nict-2*.

Variants in intergenic regions are not displayed.

position	ALT	ALT isotypes	REF seq	ALT seq	gene	amino_acid change	variant impact	consequence
3710633	CB4856	4	CATCAG	C	srx-110	226NILSPIMLPR LDLHGRRGG*	HIGH	frameshift
3718350	CB4856	1	G	A	fbxc-52	112E>112K	HIGH	missense
3721324	CB4856	387	T	C	C46E10.3.1	3D	LOW	synonymous
3721402	CB4856	381	A	G	C46E10.3.1	15L	LOW	synonymous
3721454	CB4856	382	G	A	C46E10.3.1	33V>33I	HIGH	missense
3721480	CB4856	387	A	G	C46E10.3.1	41A	LOW	synonymous
3721505	CB4856	382	G	C	C46E10.3.1	50V>50L	HIGH	missense
3721528	CB4856	387	T	C	C46E10.3.1	57F	LOW	synonymous
3721529	CB4856	382	C	G	C46E10.3.1	58P>58A	HIGH	missense
3721592	CB4856	379	G	A	C46E10.3.1	NA	HIGH	splice_region
3721592	CB4856	379	G	A	C46E10.3.1	59E	LOW	synonymous
3721610	CB4856	376	C	T	C46E10.3.1	65S	LOW	synonymous
3721616	CB4856	381	C	A	C46E10.3.1	67S	LOW	synonymous
3721633	CB4856	382	T	A	C46E10.3.1	73I>73N	HIGH	missense
3721634	CB4856	382	T	C	NA	NA	Linker	
3721635	CB4856	382	T	C	C46E10.3.1	74L	LOW	synonymous
3721642	CB4856	382	C	T	C46E10.3.1	76A>76V	HIGH	missense
3721644	CB4856	382	T	C	C46E10.3.1	77S>77P	HIGH	missense
3721649	CB4856	382	A	T	C46E10.3.1	78P	LOW	synonymous
3721758	CB4856	387	T	A	C46E10.3.1	95D>95E	HIGH	missense
3721758	CB4856	387	T	A	C46E10.3.1	NA	HIGH	splice_region
3721815	CB4856	387	A	T	C46E10.3.1	114T	LOW	synonymous
3722059	CB4856	478	C	T	C46E10.3.1	181S	LOW	synonymous
3722441	CB4856	364	C	T	C46E10.3.1	229P	LOW	synonymous
3722445	CB4856	376	T	A	C46E10.3.1	231C>231S	HIGH	missense
3722518	CB4856	377	T	C	C46E10.3.1	255L>255P	HIGH	missense
3722582	CB4856	378	C	A	C46E10.3.1	276P	LOW	synonymous
3722664	CB4856	466	A	C	C46E10.3.1	304N>304H	HIGH	missense
3722813	CB4856	370	C	T	C46E10.3.1	320S>320L	HIGH	missense
3722847	CB4856	377	T	A	C46E10.3.1	331L	LOW	synonymous
3722850	CB4856	377	G	A	C46E10.3.1	332P	LOW	synonymous
3722853	CB4856	370	T	C	C46E10.3.1	333R	LOW	synonymous
3722865	CB4856	370	T	G	C46E10.3.1	337V	LOW	synonymous
3722871	CB4856	377	A	G	C46E10.3.1	339K	LOW	synonymous
3722967	CB4856	376	A	T	C46E10.3.1	354A	LOW	synonymous
3722970	CB4856	376	T	C	C46E10.3.1	355T	LOW	synonymous
3722973	CB4856	376	C	T	C46E10.3.1	356D	LOW	synonymous
3723015	CB4856	369	C	T	C46E10.3.1	370V	LOW	synonymous
3723045	CB4856	366	G	A	C46E10.3.1	380R	LOW	synonymous
3723060	CB4856	366	CA	C	C46E10.3.1	385IKE>385IKN	HIGH	missense& inframe_altering
3723066	CB4856	366	A	AT	NA	NA	Linker	
3723074	CB4856	366	C	G	C46E10.3.1	390A>390G	HIGH	missense
3723075	CB4856	366	G	A	NA	NA	Linker	
3726322	CB4856	348	A	T	C46E10.8.1	NA	HIGH	splice_region& 3_prime_utr
3726322	CB4856	348	A	T	C46E10.8.1	NA	HIGH	splice_region& 3_prime_utr
3726398	CB4856	344	G	T	C46E10.8.1	NA	LOW	3_prime_utr
3726401	CB4856	351	G	A	C46E10.8.1	NA	LOW	3_prime_utr
3726433	CB4856	8	T	C	C46E10.8.1	NA	LOW	3_prime_utr
3726473	CB4856	348	T	G	C46E10.8.1	NA	LOW	3_prime_utr
3726481	CB4856	341	CCAAAA	C	C46E10.8.1	NA	LOW	3_prime_utr
3726496	CB4856	347	C	T	C46E10.8.1	NA	LOW	3_prime_utr
3726501	CB4856	345	G	A	C46E10.8.1	NA	LOW	3_prime_utr
3726508	CB4856	344	G	A	C46E10.8.1	NA	LOW	3_prime_utr
3726509	CB4856	343	C	A	C46E10.8.1	NA	LOW	3_prime_utr

3726512	CB4856	341	A	G	C46E10.8.1	NA	LOW	3_prime_utr
3726518	CB4856	337	C	G	C46E10.8.1	NA	LOW	3_prime_utr
3726520	CB4856	336	A	G	C46E10.8.1	NA	LOW	3_prime_utr
3726535	CB4856	309	T	C	C46E10.8.1	NA	LOW	3_prime_utr
3726541	CB4856	300	C	T	C46E10.8.1	NA	LOW	3_prime_utr
3726544	CB4856	284	C	A	C46E10.8.1	NA	LOW	3_prime_utr
3726550	CB4856	231	T	G	C46E10.8.1	NA	LOW	3_prime_utr
3726561	CB4856	257	A	T	C46E10.8.1	NA	LOW	3_prime_utr
3726593	CB4856	348	T	C	C46E10.8.1	NA	LOW	3_prime_utr
3726607	CB4856	350	G	C	C46E10.8.1	NA	LOW	3_prime_utr
3726610	CB4856	351	T	C	C46E10.8.1	NA	LOW	3_prime_utr
3726611	CB4856	351	T	C	C46E10.8.1	NA	LOW	3_prime_utr
3726615	CB4856	351	C	T	C46E10.8.1	NA	LOW	3_prime_utr
3726620	CB4856	351	C	T	C46E10.8.1	NA	LOW	3_prime_utr
3726621	CB4856	351	G	C	C46E10.8.1	NA	LOW	3_prime_utr
3726623	CB4856	351	G	A	C46E10.8.1	NA	LOW	3_prime_utr
3726626	CB4856	350	G	A	C46E10.8.1	NA	LOW	3_prime_utr
3726629	CB4856	350	T	A	C46E10.8.1	NA	LOW	3_prime_utr
3726639	CB4856	351	C	T	C46E10.8.1	NA	LOW	3_prime_utr
3726644	CB4856	351	T	C	C46E10.8.1	NA	LOW	3_prime_utr
3726669	CB4856	353	C	T	C46E10.8.1	NA	LOW	3_prime_utr
3726695	CB4856	343	T	G	C46E10.8.1	165Q>165P	HIGH	missense
3726698	CB4856	343	G	A	C46E10.8.1	164S>164L	HIGH	missense
3726815	CB4856	350	A	C	C46E10.8.1	139S	LOW	synonymous
3726821	CB4856	350	G	C	C46E10.8.1	137L	LOW	synonymous
3726828	CB4856	350	C	T	NA	NA	Linker	
3726829	CB4856	350	G	T	C46E10.8.1	135R>135N	HIGH	missense
3726834	CB4856	350	T	C	C46E10.8.1	133D>133G	HIGH	missense
3726836	CB4856	350	C	G	C46E10.8.1	132G	LOW	synonymous
3726840	CB4856	350	A	T	C46E10.8.1	131L>131H	HIGH	missense
3726888	CB4856	351	A	G	NA	NA	Linker	
3726889	CB4856	351	C	T	C46E10.8.1	115V>115T	HIGH	missense
3726941	CB4856	352	T	A	C46E10.8.1	112T>112S	HIGH	missense
3726972	CB4856	352	G	A	C46E10.8.1	101G	LOW	synonymous
3727305	CB4856	351	C	A	C46E10.8.1	28V>28F	HIGH	missense
3727399	CB4856	351	A	G	C46E10.8.1	12F	LOW	synonymous
3727456	CB4856	351	A	G	C46E10.8.1	NA	LOW	5_prime_utr
3727474	CB4856	351	C	T	C46E10.8.1	NA	LOW	5_prime_utr
3727486	CB4856	351	G	A	C46E10.8.1	NA	LOW	5_prime_utr
3727584	CB4856	351	G	A	C46E10.8.1	NA	LOW	5_prime_utr
3727604	CB4856	353	A	T	C46E10.8.1	NA	LOW	5_prime_utr
3727666	CB4856	353	A	C	C46E10.8.1	NA	LOW	5_prime_utr
3727666	CB4856	353	A	C	C46E10.8.1	NA	LOW	5_prime_utr
3727675	CB4856	351	T	A	C46E10.8.1	NA	LOW	5_prime_utr
3727702	CB4856	351	T	C	C46E10.8.1	NA	LOW	5_prime_utr
3727711	CB4856	351	G	A	C46E10.8.1	NA	LOW	5_prime_utr
3727711	CB4856	351	G	A	C46E10.8.1	NA	LOW	5_prime_utr
3727783	CB4856	351	G	A	C46E10.8.1	NA	LOW	5_prime_utr
3727810	CB4856	351	A	T	C46E10.8.1	NA	LOW	5_prime_utr
3727849	CB4856	351	G	A	C46E10.8.1	NA	LOW	5_prime_utr
3727861	CB4856	351	C	T	C46E10.8.1	NA	LOW	5_prime_utr
3727862	CB4856	353	A	G	C46E10.8.1	NA	LOW	5_prime_utr
3727962	CB4856	351	G	T	C46E10.8.1	NA	LOW	5_prime_utr
3727966	CB4856	351	A	T	C46E10.8.1	NA	LOW	5_prime_utr
3727989	CB4856	351	A	G	C46E10.8.1	NA	LOW	5_prime_utr
3727989	CB4856	351	A	G	C46E10.8.1	NA	LOW	5_prime_utr
3727992	CB4856	351	T	G	C46E10.8.1	NA	LOW	5_prime_utr
3728005	CB4856	350	GAA	G	C46E10.8.1	NA	LOW	5_prime_utr
3728011	CB4856	350	G	GGA	C46E10.8.1	NA	LOW	5_prime_utr
3728023	CB4856	353	G	A	C46E10.8.1	NA	LOW	5_prime_utr
3728068	CB4856	341	A	T	C46E10.8.1	NA	LOW	5_prime_utr
3730278	CB4856	341	A	G	NA	NA	Linker	
3730315	MY23	100	G	T	sdz-12	277T>277K	HIGH	missense
3730315	CB4856	341	G	T	NA	NA	Linker	
3730392	CB4856	341	G	A	NA	NA	Linker	
3730394	CB4856	341	T	G	NA	NA	Linker	
3730701	MY23	25	T	A	sdz-12	188T>188S	HIGH	missense
3730891	CB4856	341	T	C	NA	NA	Linker	
3730961	CB4856	341	C	T	NA	NA	Linker	

3730961	MY23	102	C	T	sdz-12	115K	LOW	synonymous
3731067	MY23	167	T	C	sdz-12	99T>99A	HIGH	missense
3731067	CB4856	341	T	C	NA	NA	Linker	
3731127	CB4856	341	T	C	NA	NA	Linker	
3731149	CB4856	341	T	C	NA	NA	Linker	
3731441	CB4856	341	G	A	NA	NA	Linker	
3731470	CB4856	330	GCA	G	sdz-12	26APVSSLQKKVCE SENAENAHEAHYV*	HIGH	frameshift
3731498	CB4856	344	G	T	sdz-12	17S	LOW	synonymous
3731579	CB4856	342	G	GGGTG GCGC	sdz-12	NA	LOW	5_prime_utr
3731688	MY23	87	A	G	sdz-12	NA	LOW	5_prime_utr
3731779	CB4856	344	G	A	sdz-12	NA	LOW	5_prime_utr
3731779	CB4856	344	G	A	sdz-12	NA	LOW	5_prime_utr
3731782	CB4856	342	T	TG	sdz-12	NA	LOW	5_prime_utr
3731787	CB4856	344	A	T	sdz-12	NA	LOW	5_prime_utr
3731809	CB4856	344	C	G	sdz-12	NA	LOW	5_prime_utr
3731820	CB4856	340	G	A	sdz-12	NA	LOW	5_prime_utr
3732086	CB4856	8	A	G	bath-31	21K>21E	HIGH	missense
3732100	CB4856	345	T	G	bath-31	25T	LOW	synonymous
3732204	CB4856	345	G	A	bath-31	45R>45K	HIGH	missense
3732292	CB4856	346	T	A	bath-31	74S	LOW	synonymous
3732300	CB4856	346	A	G	bath-31	77E>77G	HIGH	missense
3732305	CB4856	344	C	T	bath-31	79P>79S	HIGH	missense
3732319	CB4856	345	A	T	bath-31	83T	LOW	synonymous
3732327	CB4856	343	T	C	bath-31	86V>86A	HIGH	missense
3732329	CB4856	343	A	C	bath-31	87N>87H	HIGH	missense
3732423	CB4856	347	C	A	bath-31	118A>118E	HIGH	missense
3732434	CB4856	347	A	G	bath-31	122K>122E	HIGH	missense
3732664	CB4856	354	C	T	bath-31	198D	LOW	synonymous
3732676	MY23	77	C	G	bath-31	202F>202L	HIGH	missense
3732677	MY23	78	C	T	bath-31	203Q>203*	HIGH	stop_gained
3735992	CB4856	347	T	G	F12E12.6.1	85N>85H	HIGH	missense
3736023	CB4856	354	T	C	F12E12.6.1	74R	LOW	synonymous
3736044	CB4856	349	G	C	F12E12.6.1	67S	LOW	synonymous
3736050	CB4856	8	C	G	F12E12.6.1	65M>65I	HIGH	missense
3736056	CB4856	354	T	C	F12E12.6.1	63E	LOW	synonymous
3736196	CB4856	354	T	C	F12E12.6.1	33T>33A	HIGH	missense
3736217	CB4856	354	T	G	F12E12.6.1	26K>26Q	HIGH	missense
3736687	CB4856	348	T	A	F12E12.6.1	NA	LOW	5_prime_utr
3736719	CB4856	349	C	T	F12E12.6.1	NA	LOW	5_prime_utr
3739968	MY23	133	C	T	F12E12.2.1	8L	LOW	synonymous
3739970	MY23	133	A	G	NA	NA	Linker	
3739995	MY23	135	A	G	F12E12.2.1	17M>17V	HIGH	missense
3740006	MY23	135	T	C	F12E12.2.1	20D	LOW	synonymous
3740031	MY23	135	C	A	F12E12.2.1	29L>29M	HIGH	missense
3740146	MY23	135	T	C	F12E12.2.1	52S>52P	HIGH	missense
3740377	MY23	135	T	A	F12E12.2.1	77D>77E	HIGH	missense
3740439	MY23	133	C	T	F12E12.2.1	98P>98L	HIGH	missense
3740451	MY23	132	T	A	F12E12.2.1	102I>102K	HIGH	missense
3740479	MY23	132	T	C	F12E12.2.1	111D	LOW	synonymous
3740687	MY23	133	A	G	F12E12.2.1	120L	LOW	synonymous
3740695	MY23	133	T	A	F12E12.2.1	123V>123D	HIGH	missense
3740696	MY23	133	T	C	NA	NA	Linker	
3740738	MY23	133	A	G	F12E12.2.1	137K	LOW	synonymous
3740741	MY23	133	A	G	F12E12.2.1	138Q	LOW	synonymous
3740765	MY23	135	G	A	F12E12.2.1	146G	LOW	synonymous
3740790	MY23	77	T	A	F12E12.2.1	155Y>155N	HIGH	missense
3740792	MY23	77	T	C	NA	NA	Linker	
3740810	MY23	133	G	A	F12E12.2.1	161R	LOW	synonymous
3741893	MY23	85	T	C	F12E12.1.1	NA	LOW	5_prime_utr
3742010	MY23	133	T	G	F12E12.1.1	18S	LOW	synonymous
3742022	MY23	135	T	C	F12E12.1.1	22I	LOW	synonymous
3742024	MY23	133	G	A	F12E12.1.1	23R>23Q	HIGH	missense
3742034	CB4856	267	A	G	F12E12.1.1	26R	LOW	synonymous
3742163	MY23	140	G	A	F12E12.1.1	50P	LOW	synonymous
3742175	MY23	140	G	A	F12E12.1.1	54P	LOW	synonymous
3742205	MY23	133	G	A	F12E12.1.1	64E	LOW	synonymous
3742214	MY23	133	G	A	F12E12.1.1	67Q	LOW	synonymous
3742241	MY23	140	T	G	F12E12.1.1	76V	LOW	synonymous

3742265	MY23	131	T	G	F12E12.1.1	84L	LOW	synonymous
3742289	MY23	133	A	G	F12E12.1.1	92E	LOW	synonymous
3742296	MY23	133	T	C	F12E12.1.1	95L	LOW	synonymous
3742299	CB4856	376	A	C	F12E12.1.1	96K>96Q	HIGH	missense
3742299	MY23	131	A	C	F12E12.1.1	96K>96Q	HIGH	missense
3742301	MY23	131	A	G	NA	NA	Linker	
3742330	MY23	131	G	A	F12E12.1.1	106R>106K	HIGH	missense
3742355	MY23	138	T	A	F12E12.1.1	114D>114E	HIGH	missense
3742376	MY23	135	TTATTGAT CTAATTA	T	F12E12.1.1	NA	HIGH	stop_lost& inframe_deletion
3742376	MY23	135	TTATTGAT CTAATTA	T	F12E12.1.1	NA	HIGH	stop_lost& inframe_deletion
3742376	MY23	135	TTATTGAT CTAATTA	T	F12E12.1.1	NA	LOW	3_prime_utr
3742376	MY23	135	TTATTGAT CTAATTA	T	F12E12.1.1	NA	LOW	3_prime_utr
3742396	MY23	136	T	A	F12E12.1.1	NA	LOW	3_prime_utr
3742445	MY23	131	TG	T	F12E12.1.1	NA	LOW	3_prime_utr
3742447	MY23	131	ACTTTTTG	A	F12E12.1.1	NA	LOW	3_prime_utr
3742461	MY23	133	C	T	F12E12.1.1	NA	LOW	3_prime_utr
3742518	MY23	133	A	T	F12E12.11.1	NA	LOW	3_prime_utr
3742532	MY23	135	C	T	F12E12.11.1	278K	LOW	synonymous
3742576	MY23	138	G	A	F12E12.11.1	264L	LOW	synonymous
3742658	MY23	140	G	A	F12E12.11.1	236A	LOW	synonymous
3742670	MY23	132	G	A	F12E12.11.1	232P	LOW	synonymous
3742673	MY23	133	C	T	F12E12.11.1	231Q	LOW	synonymous
3742719	MY23	133	T	A	F12E12.11.1	216Y>216F	HIGH	missense
3743069	MY23	132	C	A	F12E12.11.1	190V	LOW	synonymous
3743078	MY23	133	C	T	F12E12.11.1	187V	LOW	synonymous
3743123	MY23	133	G	A	F12E12.11.1	172F	LOW	synonymous
3743135	MY23	140	A	T	F12E12.11.1	168A	LOW	synonymous
3743144	MY23	133	G	T	F12E12.11.1	165S	LOW	synonymous
3743153	MY23	133	G	A	F12E12.11.1	162Y	LOW	synonymous
3743207	MY23	132	G	A	F12E12.11.1	144I	LOW	synonymous
3743246	MY23	140	C	T	F12E12.11.1	131K	LOW	synonymous
3743257	MY23	138	A	G	F12E12.11.1	128L	LOW	synonymous
3743258	MY23	133	C	T	F12E12.11.1	127T	LOW	synonymous
3743272	MY23	133	G	T	F12E12.11.1	123R	LOW	synonymous
3743306	MY23	140	A	G	F12E12.11.1	111I	LOW	synonymous
3743345	MY23	133	T	C	F12E12.11.1	98V	LOW	synonymous
3743411	MY23	138	C	T	F12E12.11.1	76L	LOW	synonymous
3743426	MY23	134	C	T	F12E12.11.1	71Q	LOW	synonymous
3743440	MY23	130	G	A	F12E12.11.1	67L	LOW	synonymous
3743441	MY23	130	A	G	F12E12.11.1	66D	LOW	synonymous
3743474	MY23	85	T	A	F12E12.11.1	55V	LOW	synonymous
3743486	MY23	46	C	T	F12E12.11.1	51L	LOW	synonymous
3743750	MY23	123	G	A	F12E12.11.1	9A	LOW	synonymous
3743768	MY23	136	T	G	F12E12.11.1	3R	LOW	synonymous
3743802	MY23	139	T	C	F12E12.11.1	NA	LOW	5_prime_utr
3743808	MY23	139	G	GA	F12E12.11.1	NA	LOW	5_prime_utr
3746496	CB4856	344	T	C	fbxb-92	306E>308G	HIGH	missense
3746521	CB4856	348	T	A	fbxb-92	298N>300Y	HIGH	missense
3746522	CB4856	348	G	GTCATCA	fbxb-92	297Y>297YDD	HIGH	inframe_insertion
3746527	CB4856	353	T	C	fbxb-92	296N>296D	HIGH	missense
3746542	CB4856	352	T	C	fbxb-92	291T>291A	HIGH	missense
3746623	CB4856	356	T	C	fbxb-92	264I>264V	HIGH	missense
3746656	CB4856	358	A	G	fbxb-92	253Y>253H	HIGH	missense
3746657	CB4856	359	T	G	fbxb-92	252I	LOW	synonymous
3746662	CB4856	358	A	G	fbxb-92	251Y>251H	HIGH	missense
3746687	CB4856	361	C	T	fbxb-92	242G	LOW	synonymous
3746691	CB4856	360	A	T	fbxb-92	241I>241N	HIGH	missense
3746707	CB4856	357	A	G	fbxb-92	236L	LOW	synonymous
3746713	CB4856	356	C	T	fbxb-92	234V>234I	HIGH	missense
3746753	CB4856	357	A	G	fbxb-92	220I	LOW	synonymous
3746766	CB4856	355	T	C	fbxb-92	216N>216S	HIGH	missense
3746795	CB4856	356	G	A	fbxb-92	206Y	LOW	synonymous
3746804	CB4856	355	C	A	fbxb-92	203S	LOW	synonymous

3746817	CB4856	355	T	C	fbxb-92	199D>199G	HIGH	missense
3746833	CB4856	355	G	A	fbxb-92	194L	LOW	synonymous
3746946	CB4856	355	C	T	fbxb-92	156G>156E	HIGH	missense
3746967	CB4856	356	T	C	fbxb-92	149K>149R	HIGH	missense
3746987	MY23	110	T	A	fbxb-92	142L	LOW	synonymous
3746995	CB4856	363	T	C	fbxb-92	140K>140E	HIGH	missense
3747003	CB4856	363	A	T	fbxb-92	137F>137Y	HIGH	missense
3747005	CB4856	363	C	T	fbxb-92	136Q	LOW	synonymous
3747008	CB4856	363	G	A	fbxb-92	135G	LOW	synonymous
3747023	CB4856	364	T	C	NA	NA	Linker	
3747024	CB4856	362	A	G	fbxb-92	130V>130A	HIGH	missense
3747031	CB4856	363	T	G	fbxb-92	128T>128P	HIGH	missense
3747041	CB4856	363	C	T	fbxb-92	124L	LOW	synonymous
3747044	CB4856	413	C	T	fbxb-92	123A	LOW	synonymous
3747053	CB4856	365	A	T	NA	NA	Linker	
3747054	CB4856	363	C	T	fbxb-92	120R>120Q	HIGH	missense
3747077	CB4856	363	G	C	fbxb-92	112F>112L	HIGH	missense
3747080	CB4856	363	G	A	fbxb-92	111H	LOW	synonymous
3747389	CB4856	365	A	G	NA	NA	Linker	
3747390	CB4856	365	A	T	fbxb-92	24L>24T	HIGH	missense
3747396	CB4856	372	A	T	fbxb-92	22Y>22N	HIGH	missense
3747397	CB4856	366	A	G	fbxb-92	21R	LOW	synonymous
3747403	CB4856	372	C	T	fbxb-92	19V	LOW	synonymous
3747417	CB4856	366	A	G	fbxb-92	15S>15P	HIGH	missense
3747419	CB4856	372	T	C	fbxb-92	14K>14R	HIGH	missense
3747445	CB4856	371	C	A	fbxb-92	5S	LOW	synonymous
3747468	CB4856	366	A	G	fbxb-92	NA	LOW	5_prime_utr
3747478	CB4856	367	A	G	fbxb-92	NA	LOW	5_prime_utr
3747498	CB4856	367	A	G	fbxb-92	NA	LOW	5_prime_utr
3747500	CB4856	363	AT	A	fbxb-92	NA	LOW	5_prime_utr
3747749	CB4856	367	G	T	fbxb-90	NA	LOW	3_prime_utr
3747749	CB4856	367	G	T	fbxb-90	NA	LOW	3_prime_utr
3747758	CB4856	363	TG	T	fbxb-90	NA	LOW	3_prime_utr
3747760	CB4856	364	A	ATTT	fbxb-90	NA	LOW	3_prime_utr
3747761	CB4856	364	C	A	fbxb-90	NA	LOW	3_prime_utr
3747776	CB4856	371	C	A	fbxb-90	NA	LOW	3_prime_utr
3747810	CB4856	357	C	T	fbxb-90	312V>312I	HIGH	missense
3747815	CB4856	362	T	C	fbxb-90	310Q>310R	HIGH	missense
3747931	CB4856	365	A	T	fbxb-90	271H>271Q	HIGH	missense
3747934	CB4856	362	A	G	NA	NA	Linker	
3747935	CB4856	361	C	T	fbxb-90	270C>270Y	HIGH	missense
3748002	CB4856	365	A	G	fbxb-90	248F>248L	HIGH	missense
3748086	CB4856	360	G	T	fbxb-90	220Q>220K	HIGH	missense
3748120	MY23	91	T	C	fbxb-90	208L	LOW	synonymous
3748125	CB4856	372	C	T	fbxb-90	207D>207N	HIGH	missense
3748200	CB4856	365	C	A	fbxb-90	182A>182S	HIGH	missense
3748284	CB4856	367	C	T	fbxb-90	154D>154N	HIGH	missense
3748421	CB4856	366	C	T	fbxb-90	108S>108N	HIGH	missense
3748424	CB4856	366	C	T	fbxb-90	107R>107Q	HIGH	missense
3748429	CB4856	373	G	T	fbxb-90	105S	LOW	synonymous
3748448	CB4856	367	A	C	fbxb-90	99F>99C	HIGH	missense
3748461	CB4856	368	G	A	fbxb-90	95L>95F	HIGH	missense
3748499	CB4856	366	A	G	fbxb-90	82L>82S	HIGH	missense
3748508	CB4856	366	A	C	fbxb-90	79L>79R	HIGH	missense
3748543	CB4856	367	G	A	fbxb-90	67A	LOW	synonymous
3748550	CB4856	372	A	G	fbxb-90	65I>65T	HIGH	missense
3748590	CB4856	373	T	C	fbxb-90	52K>52E	HIGH	missense
3748595	CB4856	367	T	C	fbxb-90	50Q>50R	HIGH	missense
3748598	CB4856	373	A	G	fbxb-90	49F>49S	HIGH	missense
3748614	CB4856	368	T	A	fbxb-90	44T>44S	HIGH	missense
3748621	CB4856	373	C	A	fbxb-90	41L	LOW	synonymous
3748632	CB4856	373	C	T	fbxb-90	38A>38T	HIGH	missense
3748644	CB4856	368	T	C	fbxb-90	34I>34V	HIGH	missense
3748654	CB4856	343	C	A	NA	NA	Linker	
3748656	CB4856	343	G	T	fbxb-90	30P>30T	HIGH	missense
3749139	MY23	77	A	T	fbxb-91	NA	LOW	3_prime_utr
3749529	MY23	62	C	T	fbxb-91	187R>187K	HIGH	missense
3749708	MY23	25	G	A	fbxb-91	127C	LOW	synonymous
3751591	MY23	100	C	A	sdz-11	213A>213S	HIGH	missense

3752274	MY23	88	A	T	sdz-11	2V>2D	HIGH	missense
3753900	MY23	72	T	C	T24E12.5.1	NA	LOW	5_prime_utr
3753936	MY23	72	T	A	T24E12.5.1	6L>6I	HIGH	missense
3753945	MY23	109	G	A	T24E12.5.1	9A>9T	HIGH	missense
3753948	MY23	109	C	G	T24E12.5.1	10P>10A	HIGH	missense
3753962	MY23	72	AGCTGTT	A	T24E12.5.1	14EAV>14E	HIGH	inframe_deletion
3753991	MY23	72	C	T	T24E12.5.1	24S>22F	HIGH	missense
3754185	MY23	72	T	C	T24E12.5.1	89S>87P	HIGH	missense
3754192	MY23	72	G	T	T24E12.5.1	91G>89V	HIGH	missense
3754232	MY23	72	C	A	T24E12.5.1	104N>102K	HIGH	missense
3754294	MY23	72	G	A	T24E12.5.1	125C>123Y	HIGH	missense
3755610	MY23	88	G	A	T24E12.5.1	390K	LOW	synonymous
3755656	MY23	72	T	C	T24E12.5.1	406S>404P	HIGH	missense
3755672	MY23	72	A	G	T24E12.5.1	411K>409R	HIGH	missense
3755681	MY23	72	C	G	T24E12.5.1	414A>412G	HIGH	missense
3756234	MY23	72	C	A	T24E12.5.1	422N>420K	HIGH	missense
3756234	MY23	86	C	A	T24E12.5.1	NA	HIGH	splice_region
3756243	MY23	72	T	A	T24E12.5.1	425R	LOW	synonymous
3756344	MY23	72	G	A	T24E12.5.1	459R>457Q	HIGH	missense
3756409	MY23	72	A	G	T24E12.5.1	481T>479A	HIGH	missense
3756411	MY23	90	A	G	NA	NA	Linker	
3756414	MY23	97	C	T	T24E12.5.1	482G	LOW	synonymous
3756416	MY23	72	G	A	T24E12.5.1	483R>481Q	HIGH	missense
3756468	MY23	98	A	G	T24E12.5.1	500P	LOW	synonymous
3756479	MY23	72	C	A	T24E12.5.1	504T>502K	HIGH	missense
3756835	MY23	76	C	T	T24E12.5.1	523I	LOW	synonymous
3756858	MY23	72	G	A	T24E12.5.1	531G>529D	HIGH	missense
3756912	MY23	71	TAC	T	T24E12.5.1	547MPID*	HIGH	frameshift& stop_retained
3756925	MY23	75	A	T	NA	NA	Linker	
3756928	MY23	75	A	T	NA	NA	Linker	
3756959	MY23	75	A	T	NA	NA	Linker	
3756961	MY23	75	T	C	NA	NA	Linker	
3756968	MY23	75	A	C	NA	NA	Linker	
3756994	MY23	75	G	C	NA	NA	Linker	
3757052	MY23	75	C	T	NA	NA	Linker	
3757067	MY23	75	A	G	NA	NA	Linker	
3757069	MY23	75	A	G	NA	NA	Linker	
3757102	MY23	75	G	C	NA	NA	Linker	
3757171	MY23	75	A	G	NA	NA	Linker	
3757204	MY23	75	G	C	NA	NA	Linker	
3757247	MY23	71	A	G	NA	NA	Linker	
3757269	MY23	75	T	C	NA	NA	Linker	
3757283	MY23	75	A	C	NA	NA	Linker	
3757291	MY23	75	A	G	NA	NA	Linker	
3757292	MY23	75	T	C	NA	NA	Linker	
3757294	MY23	75	A	C	NA	NA	Linker	
3757332	MY23	75	T	C	NA	NA	Linker	
3757334	MY23	75	C	A	NA	NA	Linker	
3757335	MY23	75	C	G	NA	NA	Linker	
3757346	MY23	76	A	G	NA	NA	Linker	
3757362	MY23	76	C	T	T24E12.5.1	NA	LOW	3_prime_utr
3757368	MY23	84	C	T	T24E12.5.1	NA	LOW	3_prime_utr
3757815	CB4856	373	A	T	srx-111	6Q>6L	HIGH	missense
3757963	CB4856	365	G	A	srx-111	19A>19T	HIGH	missense
3757963	CB4856	365	G	A	srx-111	NA	HIGH	splice_region
3757986	CB4856	365	C	T	srx-111	26F	LOW	synonymous
3757995	CB4856	365	C	T	srx-111	29S	LOW	synonymous
3758016	CB4856	369	A	T	srx-111	36S	LOW	synonymous
3758085	CB4856	366	T	G	srx-111	59V	LOW	synonymous
3758088	CB4856	366	T	C	srx-111	60Y	LOW	synonymous
3758089	CB4856	366	T	C	srx-111	61L	LOW	synonymous
3758092	CB4856	366	T	C	srx-111	62L	LOW	synonymous
3758103	CB4856	366	C	A	srx-111	65G	LOW	synonymous
3758115	MY23	25	G	A	srx-111	NA	HIGH	splice_region
3758115	MY23	25	G	A	srx-111	69L	LOW	synonymous
3758116	CB4856	366	T	C	srx-111	NA	HIGH	splice_region
3758116	CB4856	366	T	C	srx-111	70L	LOW	synonymous

3758169	CB4856	366	G	A	srx-111	72A>72T	HIGH	missense
3758177	CB4856	370	G	T	srx-111	74L	LOW	synonymous
3758183	CB4856	367	C	A	srx-111	76P	LOW	synonymous
3758204	CB4856	371	T	C	srx-111	83L	LOW	synonymous
3758207	CB4856	371	C	T	srx-111	84N	LOW	synonymous
3758213	CB4856	365	A	G	srx-111	86A	LOW	synonymous
3758218	CB4856	365	C	T	srx-111	88A>88V	HIGH	missense
3758226	CB4856	372	C	T	srx-111	91L	LOW	synonymous
3758228	CB4856	372	A	G	NA	NA	Linker	
3758949	CB4856	360	A	G	srx-111	225I>225V	HIGH	missense
3758949	MY23	78	A	G	srx-111	225I>225V	HIGH	missense
3758951	CB4856	360	C	T	NA	NA	Linker	
3759002	MY23	25	A	T	srx-111	242L>242F	HIGH	missense
3759559	MY23	141	AT	A	T24E12.12.1	NA	LOW	5_prime_utr
3759654	MY23	66	T	C	T24E12.12.1	7H	LOW	synonymous
3760307	CB4856	229	C	T	T24E12.6b.1	330K	LOW	synonymous
3760307	CB4856	229	C	T	T24E12.6a.1	330K	LOW	synonymous
3760979	MY23	72	G	A	T24E12.6b.1	15Y	LOW	synonymous
3760979	MY23	72	G	A	T24E12.6a.1	15Y	LOW	synonymous
3760979	MY23	72	G	A	T24E12.6b.1	193Y	LOW	synonymous
3760979	MY23	72	G	A	T24E12.6a.1	193Y	LOW	synonymous
3764481	CB4856	8	G	A	srx-112	20F	LOW	synonymous
3765201	CB4856	224	T	A	T24E12.10.1	NA	LOW	3_prime_utr
3765918	CB4856	219	G	A	T24E12.10.1	1176P>1176L	HIGH	missense
3766060	MY23	25	G	A	T24E12.10.1	1144A>1144V	HIGH	missense
3766321	MY23	25	C	T	T24E12.10.1	1094G>1094S	HIGH	missense
3766352	CB4856	221	C	T	T24E12.10.1	1083Q	LOW	synonymous
3768292	CB4856	222	A	C	T24E12.10.1	715M>715R	HIGH	missense
3768531	CB4856	222	C	T	T24E12.10.1	650D>650N	HIGH	missense
3771881	MY23	34	C	T	T24E12.10.1	243D>243N	HIGH	missense
3772589	CB4856	225	C	T	T24E12.10.1	92K	LOW	synonymous
3783872	CB4856	4	A	C	T24E12.2.1	172Q>172P	HIGH	missense
3784129	MY23	25	G	A	T24E12.2.1	243G	LOW	synonymous
3791495	MY23	20	G	A	Y8A9A.4.1	NA	LOW	3_prime_utr
3791495	MY23	20	G	A	Y8A9A.3.1	NA	LOW	5_prime_utr
3791915	CB4856	1	T	G	Y8A9A.3.1	112V	LOW	synonymous
3797457	MY23	8	G	A	Y8A9A.2.1	544V>544I	HIGH	missense

Table 2. Single-cell RNA-seq data of *nta-1* and *nep-2*.

Top 20 cell types for expression of *nta-1* and *nep-2*. Glia-1 and Glia-2 may represent GLR cells. The data was obtained from CeNGEN (Taylor et al., 2021).

	nta-1			***nep-2***	
	Cell type	**Expression level**		**Cell type**	**Expression level**
1	Glia_1	199.999	1	PVN	495.177
2	Pharyngeal_gland_cell	180.375	2	RIR	249.069
3	Head mesodermal cells	175.185	3	LUA	233.892
4	Glia_2	161.013	4	AIM	228.899
5	Coelomocyte	121.687	5	AVM	196.734
6	PHso	84.819	6	Glia_1	187.049
7	CEPsh	61.34	7	SMD	174.148
8	ALA	57.557	8	RID	159.898
9	Marginal_cell	43.195	9	SIA	144.26
10	Pharyngeal_muscle	40.361	10	Glia_2	142.156
11	URX	30.486	11	SIB	111.378
12	ASG	29.717	12	PVP	85.383
13	Intestine	24.482	13	PVM	75.616
14	BAG	20.391	14	ADA	74.911
15	Glia_3	16.315	15	PHC	73.648
16	Body_wall_muscle	15.016	16	Uterine_cell	57.405
17	Gonadal_sheath_cell	14.573	17	DVC	57.234
18	Arcade_cell	10.831	18	DVB	48.507
19	Germline	6.237	19	ALA	44.465
20	Body_wall_muscle_anterior	5.712	20	PVD	40.508

Table 3. BLASTP hits for *nta-1* in several model organisms.

The list of proteins searched by BlastP of *nta-1*. BlastP was performed in NCBI using landmark database except *C. elegans*. Only lists with a total score above 100 are displayed.

Scientific Name	Description	Total Score	Query Cover	E value	Per. Ident	Acc. Len
Drosophila melanogaster	uncharacterized protein Dmel_CG12171	160	93%	1.00E-46	39.39%	257
Drosophila melanogaster	uncharacterized protein Dmel_CG3699	157	92%	1.00E-45	40.23%	251
Drosophila melanogaster	uncharacterized protein Dmel_CG31549, isoform B	150	93%	1.00E-42	38.31%	257
Danio rerio	uncharacterized protein LOC449555	149	92%	1.00E-42	36.92%	265
Drosophila melanogaster	uncharacterized protein Dmel_CG31548	146	92%	3.00E-41	37.21%	256
Streptomyces	SDR family oxidoreductase	139	93%	2.00E-38	36.53%	255
Bacillus	glucose 1-dehydrogenase	133	93%	1.00E-36	36.64%	248
Thermotoga maritima	3-oxoacyl-[acyl-carrier-protein] reductase	132	92%	4.00E-36	36.23%	246
Pseudomonas	SDR family oxidoreductase	132	94%	5.00E-36	35.71%	253
Glycine max	Glucose and ribitol dehydrogenase-like	130	92%	1.00E-34	35.74%	294
Bacillus	SDR family oxidoreductase	130	93%	5.00E-35	34.22%	273
Bacillus	SDR family oxidoreductase	129	92%	2.00E-34	34.73%	286
Bacillus	SDR family oxidoreductase	129	92%	2.00E-34	36.92%	289
Pseudomonas	SDR family oxidoreductase	128	93%	5.00E-34	35.88%	286
Schizosaccharomyces pombe	putative 3-hydroxyacyl-CoA dehydrogenase	127	92%	1.00E-33	33.83%	286
Thermotoga	SDR family oxidoreductase	127	92%	3.00E-34	37.40%	251
Bacillus	SDR family oxidoreductase	127	92%	1.00E-33	34.87%	285
Glycine max	seed maturation protein PM34	125	92%	5.00E-33	32.96%	293
Arabidopsis thaliana	NAD(P)-binding Rossmann-fold superfamily protein	124	92%	3.00E-32	32.33%	335
Arabidopsis thaliana	NAD(P)-binding Rossmann-fold superfamily protein	124	92%	4.00E-32	32.33%	337
Mus musculus	17-beta-hydroxysteroid dehydrogenase 14	122	96%	5.00E-32	30.91%	273
Danio rerio	dehydrogenase/reductase SDR family member 4	120	92%	4.00E-31	33.98%	276
Glycine max	glucose and ribitol dehydrogenase homolog 1-like	120	92%	4.00E-31	31.95%	293
Mycobacterium tuberculosis complex	SDR family oxidoreductase	119	92%	4.00E-31	31.94%	255

Danio rerio	uncharacterized protein LOC449555 isoform X1	119	62%	1.00E-31	39.77%	208
Deinococcus radiodurans	SDR family oxidoreductase	118	92%	5.00E-30	33.20%	304
Enterobacteriaceae	SDR family oxidoreductase	118	94%	2.00E-30	31.84%	263
Enterobacteriaceae	SDR family oxidoreductase	118	94%	2.00E-30	31.84%	263
Danio rerio	uncharacterized protein LOC541322	118	93%	9.00E-29	31.20%	551
Danio rerio	uncharacterized protein LOC541322 isoform X1	118	93%	4.00E-29	31.20%	447
Synechocystis	3-oxoacyl-[acyl-carrier-protein] reductase	117	93%	2.00E-30	33.46%	247
Caenorhabditis elegans	Dehydrogenase/reductase SDR family member 4	117	92%	5.00E-30	31.54%	260
Bacillus	dihydroanticapsin 7-dehydrogenase	117	93%	2.00E-30	31.06%	253
Arabidopsis thaliana	NAD(P)-binding Rossmann-fold superfamily protein	117	92%	5.00E-30	33.33%	289
Glycine max	NADPH-dependent aldehyde reductase 1, chloroplastic	116	92%	2.00E-29	32.45%	294
Microcystis	3-oxoacyl-[acyl-carrier-protein] reductase	115	91%	2.00E-29	31.52%	258
Sulfolobus acidocaldarius	SDR family oxidoreductase	115	95%	2.00E-29	35.79%	252
Drosophila melanogaster	uncharacterized protein Dmel_CG31546	115	93%	2.00E-29	35.36%	264
Deinococcus radiodurans	3-oxoacyl-[acyl-carrier-protein] reductase	114	91%	3.00E-29	33.85%	254
Bacillus	glucose 1-dehydrogenase	114	95%	4.00E-29	28.73%	261
Streptomyces	SDR family oxidoreductase	114	92%	4.00E-29	33.85%	245
Glycine max	short-chain dehydrogenase reductase 2a	114	92%	2.00E-28	30.69%	298
Clostridioides difficile	3-oxoacyl-[acyl-carrier-protein] reductase	112	93%	3.00E-28	33.08%	249
Glycine max	short-chain dehydrogenase reductase 2a	112	92%	9.00E-28	30.36%	341
Glycine max	uncharacterized protein LOC100306108	112	92%	4.00E-28	33.46%	266
Enterobacteriaceae	2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase EntA	111	93%	5.00E-28	34.96%	248
Glycine max	short-chain dehydrogenase reductase 3b	111	94%	1.00E-27	30.88%	267
Streptomyces	glucose 1-dehydrogenase	110	93%	1.00E-27	33.46%	255
Arabidopsis thaliana	NAD(P)-binding Rossmann-fold superfamily protein	110	92%	1.00E-27	33.33%	264
Microcystis	SDR family oxidoreductase	110	91%	1.00E-27	33.33%	266
Enterobacteriaceae	2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase EntA	109	93%	2.00E-27	34.96%	248
Bacillus	enoyl-[acyl-carrier-protein] reductase FabL	109	92%	3.00E-27	32.44%	250
Arabidopsis thaliana	NAD(P)-binding Rossmann-fold superfamily protein	109	92%	1.00E-26	30.69%	303
Clostridioides difficile	SDR family NAD(P)-dependent oxidoreductase	108	94%	9.00E-27	32.21%	262
Glycine max	(-)-isopiperitenol/(-)-carveol dehydrogenase, mitochondrial isoform X1	107	93%	4.00E-26	27.17%	286

Glycine max	(-)-isopiperitenol/(-)-carveol dehydrogenase, mitochondrial isoform X2	107	93%	4.00E-26	27.17%	283
Mus musculus	dehydrogenase/reductase SDR family member 6 isoform 1	107	94%	2.00E-26	31.20%	255
Mus musculus	dehydrogenase/reductase SDR family member 6 isoform 2	107	94%	1.00E-26	31.20%	245
Danio rerio	dehydrogenase/reductase SDR family member 6	107	94%	2.00E-26	28.95%	245
Escherichia coli	SDR family oxidoreductase	107	92%	1.00E-26	30.68%	253
Bacillus	3-oxoacyl-[acyl-carrier-protein] reductase	106	92%	4.00E-26	30.38%	246
Glycine max	xanthoxin dehydrogenase	106	92%	8.00E-26	34.80%	280
Enterobacteriaceae	SDR family oxidoreductase	105	92%	1.00E-25	30.30%	253
Streptomyces	SDR family oxidoreductase	105	93%	1.00E-25	32.58%	250
Streptomyces	SDR family oxidoreductase	105	94%	1.00E-25	31.97%	253
Glycine max	momilactone A synthase-like	104	92%	4.00E-25	31.32%	269
Deinococcus radiodurans	SDR family oxidoreductase	104	92%	3.00E-25	33.08%	258
Bacillus	SDR family oxidoreductase	104	92%	5.00E-25	33.08%	299
Dictyostelium discoideum AX4	short-chain dehydrogenase/reductase family protein	104	92%	4.00E-25	30.04%	281
Glycine max	xanthoxin dehydrogenase	104	92%	3.00E-25	34.19%	280
Pseudomonas	3-oxoacyl-ACP reductase RhlG	103	93%	4.00E-25	31.20%	256
Thermotoga	glucose 1-dehydrogenase	103	93%	4.00E-25	29.92%	256
Arabidopsis thaliana	NAD(P)-binding Rossmann-fold superfamily protein	103	96%	1.00E-24	28.94%	322
Clostridioides difficile	SDR family oxidoreductase	103	91%	3.00E-25	30.08%	245
Bacillus	SDR family oxidoreductase	103	93%	6.00E-25	28.09%	258
Pseudomonas	SDR family oxidoreductase	102	92%	1.00E-24	32.71%	255
Pseudomonas	SDR family oxidoreductase	102	91%	1.00E-24	34.10%	248
Glycine max	short-chain dehydrogenase reductase 3b-like	102	93%	1.00E-24	30.63%	264
Glycine max	tropinone reductase	102	92%	1.00E-24	33.71%	267
Homo sapiens	(3R)-3-hydroxyacyl-CoA dehydrogenase	101	92%	4.00E-24	32.09%	261

Table 4. List of nematode orthologs of *nta-1*.

This table, obtained from Wormbase, shows *nta-1* orthologs in nematode species.

Species	Ortholog	Method	Species	Ortholog	Method
C. angaria	Cang_2012_03_13_00477.g11299	WormBase-Compara	*C. quiockensis*	CSP38.g2756	WormBase-Compara
C. angaria	Cang_2012_03_13_00180.g6344	WormBase-Compara	*C. quiockensis*	CSP38.g19982	WormBase-Compara
C. angaria	Cang_2012_03_13_00477.g11300	WormBase-Compara	*C. quiockensis*	CSP38.g4637	WormBase-Compara
C. angaria	Cang_2012_03_13_00180.g6346	WormBase-Compara	*C. quiockensis*	CSP38.g19983	WormBase-Compara
C. bovis	CBOVI.g1371	WormBase-Compara	*C. quiockensis*	CSP38.g15402	WormBase-Compara
C. brenneri	CBN32397	WormBase-Compara	*C. quiockensis*	CSP38.g4633	WormBase-Compara
C. brenneri	CBN30598	WormBase-Compara	*C. remanei*	CRE09178	WormBase-Compara
C. brenneri	CBN06386	WormBase-Compara	*C. remanei*	CRE18547	WormBase-Compara
C. brenneri	CBN05915	WormBase-Compara	*C. remanei*	GCK72_019244	WormBase-Compara
C. brenneri	CBN24429	WormBase-Compara	*C. sinica*	Csp5_scaffold_00294.g8853	WormBase-Compara
C. brenneri	CBN07012	WormBase-Compara	*C. sinica*	Csp5_scaffold_00294.g8852	WormBase-Compara
C. brenneri	CBN00534	WormBase-Compara	*C. sinica*	Csp5_scaffold_00294.g8851	WormBase-Compara
C. brenneri	CBN15781	WormBase-Compara	*C. tribulationis*	CSP40.g15589	WormBase-Compara
C. brenneri	CBN32994	WormBase-Compara	*C. tribulationis*	CSP40.g15588	WormBase-Compara
C. brenneri	CBN22283	WormBase-Compara	*C. tribulationis*	CSP40.g15590	WormBase-Compara
C. briggsae	CBG06323	WormBase-Compara	*C. tropicalis*	Csp11.Scaffold629.g7621	WormBase-Compara
C. briggsae	CBG30900	WormBase-Compara	*C. tropicalis*	Csp11.Scaffold629.g7623	WormBase-Compara
C. briggsae	CBG06324	WormBase-Compara	*C. tropicalis*	Csp11.Scaffold629.g7626	WormBase-Compara
C. briggsae	CBG30901	WormBase-Compara	*C. tropicalis*	Csp11.Scaffold629.g7622	WormBase-Compara
C. elegans	chrII_pilon.g4091	WormBase-Compara	*C. uteleia*	CSP31.g13480	WormBase-Compara
C. inopinata	Sp34_10265700	WormBase-Compara	*C. uteleia*	CSP31.g13481	WormBase-Compara
C. inopinata	Sp34_50064800	WormBase-Compara	*C. uteleia*	CSP31.g13475	WormBase-Compara
C. japonica	CJA35151	TreeFam	*C. uteleia*	CSP31.g13479	WormBase-Compara
		WormBase-Compara	*C. uteleia*	CSP31.g13477	WormBase-Compara
C. japonica	CJA01047	TreeFam	*C. uteleia*	CSP31.g13476	WormBase-Compara
		WormBase-Compara	*C. uteleia*	CSP31.g13478	WormBase-Compara
C. japonica	CJA09557	TreeFam	*C. zanzibari*	CSP26.g8253	WormBase-Compara
C. nigoni	Cnig_chr_V.g18187	WormBase-Compara	*C. zanzibari*	CSP26.g8252	WormBase-Compara
C. nigoni	Cnig_chr_V.g18188	WormBase-Compara	*C. zanzibari*	CSP26.g8254	WormBase-Compara
C. quiockensis	CSP38.g19979	WormBase-Compara	*O. tipulae*	OTIPU.nOt.2.0.1.g12168	WormBase-Compara
C. quiockensis	CSP38.g17495	WormBase-Compara	*O. tipulae*	OTIPU.nOt.2.0.1.g08794	WormBase-Compara
C. quiockensis	CSP38.g4636	WormBase-Compara	*P. pacificus*	PPA24195	Inparanoid_8
C. quiockensis	CSP38.g7804	WormBase-Compara	*P. pacificus*	PPA27598	Inparanoid_8
C. quiockensis	CSP38.g4635	WormBase-Compara	*P. pacificus*	PPA14265	Inparanoid_8
C. quiockensis	CSP38.g4634	WormBase-Compara	*P. pacificus*	PPA17409	Inparanoid_8
C. quiockensis	CSP38.g19981	WormBase-Compara	*P. pacificus*	PPA17402	Inparanoid_8
C. quiockensis	CSP38.g19984	WormBase-Compara	*P. pacificus*	PPA06264	Inparanoid_8
C. quiockensis	CSP38.g2756	WormBase-Compara	*T. muris*	TMUE_3000012689	WormBase-Compara
C. quiockensis	CSP38.g19982	WormBase-Compara	*T. muris*	TMUE_0000000641	WormBase-Compara
C. quiockensis	CSP38.g4637	WormBase-Compara	*T. muris*	TMUE_2000007011	WormBase-Compara

Table 5. Summary of the distributions of *nta-1* promoter haplotypes.

*Others: 4 isotypes (3: CB, 1: MY) without collection information are included.

Isotype					
Region / *nta-1p* haplotype	Global	Hawaii	Europe	American west	Others*
CB4856	409	57	179	26	147
MY23	133	97	4	25	7
Others	8	8	0	0	0
Total	550	162	183	51	154
Frequency of MY23 haplotype	0.24	0.60	0.02	0.49	0.05

Table 6. Distributions of *nta-1* promoter haplotypes among 4 Hawaiian *C. elegans* populations.

Four Hawaiian population was defined in Crombie et al. (Crombie et al., 2019).

(a) 43 isotypes divided previously in Crombie et al.

nta-1p haplotype / Hawaii population	CB4856	MY23
Hi_Divergent	5	4
Volcano	0	8
Hi_Low	7	1
Hi_Invaded	3	15

(b) Reconstructed groups based on the latest genome-wide tree of 550 isotypes.

nta-1p haplotype / Hawaii population	CB4856	MY23	others
Hi_Divergent	22	25	8
Volcano	0	46	0
Hi_Low_1	18	1	0
Hi_Low_2	3	3	0
Hi_Invaded_1	0	9	0
Hi_Invaded_2	10	16	0

References

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, *25*(17), 3389-3402.

Altun, Z., & Hall, D. (2009). Muscle system, GLR cells. *Worm Atlas*, *10*.

Altun, Z. F., Chen, B., Wang, Z. W., & Hall, D. H. (2009). High resolution map of Caenorhabditis elegans gap junction proteins. *Developmental dynamics: an official publication of the American Association of Anatomists*, *238*(8), 1936-1950.

Andersen, E. C., Gerke, J. P., Shapiro, J. A., Crissman, J. R., Ghosh, R., Bloom, J. S., Félix, M.-A., & Kruglyak, L. (2012). Chromosome-scale selective sweeps shape Caenorhabditis elegans genomic diversity. *Nature genetics*, *44*(3), 285-290.

Andersen, E. C., & Rockman, M. V. (2022). Natural genetic variation as a tool for discovery in Caenorhabditis nematodes. *Genetics*, *220*(1), iyab156.

Antebi, A. (2015). Nuclear receptor signal transduction in C. elegans. *WormBook*, *1*(10.1895).

Arribere, J. A., Bell, R. T., Fu, B. X., Artiles, K. L., Hartman, P. S., & Fire, A. Z. (2014). Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in Caenorhabditis elegans. *Genetics*, *198*(3), 837-846.

Barriere, A., & Felix, M. A. (2005). High local genetic diversity and low outcrossing rate in Caenorhabditis elegans natural populations. *Curr Biol*, *15*(13), 1176-1184.

Beets, I., Zhang, G., Fenk, L. A., Chen, C., Nelson, G. M., Félix, M.-A., & de Bono, M. (2020). Natural variation in a dendritic scaffold protein remodels experience-dependent plasticity by altering neuropeptide expression. *Neuron*, *105*(1), 106-121. e110.

Bendesky, A., & Bargmann, C. I. (2011). Genetic contributions to behavioural diversity at the gene-environment interface. *Nat Rev Genet*, *12*(12), 809-820.

Bendesky, A., Tsunozaki, M., Rockman, M. V., Kruglyak, L., & Bargmann, C. I. (2011). Catecholamine receptor polymorphisms affect decision-making in C. elegans. *Nature*, *472*(7343), 313-318.

Benzer, S. (1971). From the gene to behavior. *Jama*, *218*(7), 1015-1022.

Brenner, S. (1974). The genetics of Caenorhabditis elegans. *Genetics*, *77*(1), 71-94.

Brown, F. D., D'Anna, I., & Sommer, R. J. (2011). Host-finding behaviour in the nematode Pristionchus pacificus. *Proc Biol Sci*, *278*(1722), 3260-3269.

Butcher, R. A., Fujita, M., Schroeder, F. C., & Clardy, J. (2007). Small-molecule pheromones that control dauer development in Caenorhabditis elegans. *Nat Chem Biol*, *3*(7), 420-422.

Campbell, J. F., & Gaugler, R. (1993). Nictation Behavior and Its Ecological Implications in the Host Search Strategies of Entomopathogenic Nematodes (Heterorhabditidae and Steinernematidae). *Behaviour*, *126*, 155-169.

Cassada, R. C., & Russell, R. L. (1975). The dauerlarva, a post-embryonic developmental variant of the nematode Caenorhabditis elegans. *Dev Biol*, *46*(2), 326-342.

Castelletto, M. L., Gang, S. S., Okubo, R. P., Tselikova, A. A., Nolan, T. J., Platzer, E. G., Lok, J. B., & Hallem, E. A. (2014). Diverse host-seeking behaviors of skin-penetrating nematodes. *PLoS Pathog*, *10*(8), e1004305.

Cockx, B., Van Bael, S., Boelen, R., Vandewyer, E., Yang, H., Le, T. A., Dalzell, J. J., Beets, I., Ludwig, C., & Lee, J. (2023). Mass Spectrometry–Driven Discovery of Neuropeptides Mediating Nictation Behavior of Nematodes. *Molecular & Cellular Proteomics*, *22*(2).

Consortium, C. e. S. (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. *science*, *282*(5396), 2012-2018.

Consortium, G. P. (2010). A map of human genome variation from population scale sequencing. *Nature*, *467*(7319), 1061.

Cook, D. E., Zdraljevic, S., Roberts, J. P., & Andersen, E. C. (2017). CeNDR, the Caenorhabditis elegans natural diversity resource. *Nucleic Acids Res*, *45*(D1), D650-D657.

Crombie, T. A., Zdraljevic, S., Cook, D. E., Tanny, R. E., Brady, S. C., Wang, Y., Evans, K. S., Hahnel, S., Lee, D., & Rodriguez, B. C. (2019). Deep sampling of Hawaiian Caenorhabditis elegans reveals high genetic diversity and admixture with global populations. *Elife*, *8*, e50465.

Darwin, C. (1882). On the dispersal of freshwater bivalves. *Nature*, *25*(649), 529-530.

Davis, P., Zarowiecki, M., Arnaboldi, V., Becerra, A., Cain, S., Chan, J., Chen, W. J., Cho, J., da Veiga Beltrame, E., & Diamantakis, S. (2022). WormBase in 2022—data, processes, and tools for analyzing Caenorhabditis elegans. *Genetics*, *220*(4), iyac003.

Desnoyers, S., Blanchard, P.-G., St-Laurent, J.-F., Gagnon, S. N., & Baillie, D. L. (2007). Caenorhabditis elegans LET-767 is able to metabolize androgens and estrogens and likely shares common ancestor with human types 3 and 12 17β-hydroxysteroid dehydrogenases. *Journal of Endocrinology*, *195*(2), 271-279.

Felix, M. A., & Braendle, C. (2010). The natural history of Caenorhabditis elegans. *Curr Biol*, *20*(22), R965-969.

Fielenbach, N., & Antebi, A. (2008). C. elegans dauer formation and the molecular basis of plasticity. *Genes & development*, *22*(16), 2149-2165.

Frézal, L., & Felix, M.-A. (2015). The natural history of model organisms: C. elegans outside the Petri dish. *Elife*, *4*, e05849.

Gans, C., & Burr, A. H. J. (1994). Unique locomotory mechanism of Mermis nigrescens, a large nematode that crawls over soil and climbs through vegetation. *J Morphol*, *222*(2), 133-148.

Gasch, A. P., Payseur, B. A., & Pool, J. E. (2016). The power of natural variation for model organism biology. *TRENDS in Genetics*, *32*(3), 147-154.

Golden, J. W., & Riddle, D. L. (1984). The Caenorhabditis elegans dauer larva: developmental effects of pheromone, food, and temperature. *Dev Biol*, *102*(2), 368-378.

Granzer, M., & Haas, W. (1991). Host-finding and host recognition of infective Ancylostoma caninum larvae. *Int J Parasitol*, *21*(4), 429-440.

Greene, J. S., Brown, M., Dobosiewicz, M., Ishida, I. G., Macosko, E. Z., Zhang, X., Butcher, R. A., Cline, D. J., McGrath, P. T., & Bargmann, C. I. (2016). Balancing selection shapes density-dependent foraging behaviour. *Nature*, *539*(7628), 254-258.

Greene, J. S., Dobosiewicz, M., Butcher, R. A., McGrath, P. T., & Bargmann, C. I. (2016). Regulatory changes in two chemoreceptor genes contribute to a Caenorhabditis elegans QTL for foraging behavior. *Elife*, *5*, e21454.

Harvey, S. C., Barker, G. L., Shorto, A., & Viney, M. E. (2009). Natural variation in gene expression in the early development of dauer larvae of Caenorhabditis elegans. *BMC genomics*, *10*, 1-11.

Harvey, S. C., Shorto, A., & Viney, M. E. (2008). Quantitative genetic analysis of life-history traits of Caenorhabditis elegans in stressful environments. *BMC evolutionary biology*, *8*(1), 1-16.

Hernandez, A. D., & Sukhdeo, M. V. (1995). Host grooming and the transmission strategy of Heligmosomoides polygyrus. *J Parasitol*, *81*(6), 865-869.

Hervé, J., Pluciennik, F., Verrecchia, F., Bastide, B., Delage, B., Joffre, M., & Deleze, J. (1996). Influence of the molecular structure of steroids on their ability to interrupt gap junctional communication. *The Journal of membrane biology*, *149*, 179-187.

Ishibashi, N., & Kondo, E. (1990). Behavior of infective juveniles. In *Entomopathogenic nematodes in biological control* (pp. 139-150). CRC Press Boca Raton.

Jeong, P. Y., Jung, M., Yim, Y. H., Kim, H., Park, M., Hong, E., Lee, W., Kim, Y. H., Kim, K., & Paik, Y. K. (2005). Chemical structure and biological activity of the Caenorhabditis elegans dauer-inducing pheromone. *Nature*, *433*(7025), 541-545.

Kiontke, K., & Sudhaus, W. (2006). Ecology of Caenorhabditis species. *WormBook*, *9*, 1-14.

Klass, M., & Hirsh, D. (1976). Non-ageing developmental variant of Caenorhabditis elegans. *Nature*, *260*(5551), 523-525.

Koenig, D., Hagmann, J., Li, R., Bemm, F., Slotte, T., Neuffer, B., Wright, S. I., & Weigel, D. (2019). Long-term balancing selection drives evolution of immunity genes in Capsella. *Elife*, *8*, e43606.

Konopka, R. J., & Benzer, S. (1971). Clock mutants of Drosophila melanogaster. *Proceedings of the National Academy of Sciences*, *68*(9), 2112-2116.

Lee, D., Lee, H., Kim, N., Lim, D. S., & Lee, J. (2017). Regulation of a hitchhiking behavior by neuronal insulin and TGF-beta signaling in the nematode Caenorhabditis elegans. *Biochem Biophys Res Commun*, *484*(2), 323-330.

Lee, D., Yang, H., Kim, J., Brady, S., Zdraljevic, S., Zamanian, M., Kim, H., Paik, Y. K., Kruglyak, L., Andersen, E. C., & Lee, J. (2017). The genetic basis of natural variation in a phoretic behavior. *Nat Commun*, *8*(1), 273.

Lee, D., Zdraljevic, S., Stevens, L., Wang, Y., Tanny, R. E., Crombie, T. A., Cook, D. E., Webster, A. K., Chirakar, R., & Baugh, L. R. (2021). Balancing selection maintains hyper-divergent haplotypes in Caenorhabditis elegans. *Nature ecology & evolution*, *5*(6), 794-807.

Lee, H., Choi, M. K., Lee, D., Kim, H. S., Hwang, H., Kim, H., Park, S., Paik, Y. K., & Lee, J. (2011). Nictation, a dispersal behavior of the nematode Caenorhabditis elegans, is regulated by IL2 neurons. *Nat Neurosci*, *15*(1), 107-112.

Lee, J. S., Shih, P. Y., Schaedel, O. N., Quintero-Cadena, P., Rogers, A. K., & Sternberg, P. W. (2017). FMRFamide-like peptides expand the behavioral repertoire of a densely connected nervous system. *Proc Natl Acad Sci U S A*, *114*(50), E10726-E10735.

Lopez, L. C. S., Rodrigues, P. J. F. P., & Rios, R. I. (1999). Frogs and snakes as phoretic dispersal agents of bromeliad ostracods (Limnocytheridae : Elpidium) and annelids (Naididae : Dero). *Biotropica*, *31*(4), 705-708.

Lösel, R., & Wehling, M. (2003). Nongenomic actions of steroid hormones. *Nature reviews Molecular cell biology*, *4*(1), 46-55.

Lukacik, P., Keller, B., Bunkoczi, G., Kavanagh, K., Hwa Lee, W., Adamski, J., & Oppermann, U. (2007). Structural and biochemical characterization of human orphan DHRS10 reveals a novel cytosolic enzyme with steroid dehydrogenase activity. *Biochemical Journal*, *402*(3), 419-427.

Macosko, E. Z., Pokala, N., Feinberg, E. H., Chalasani, S. H., Butcher, R. A., Clardy, J., & Bargmann, C. I. (2009). A hub-and-spoke circuit drives pheromone attraction and social behaviour in C. elegans. *Nature*, *458*(7242), 1171-1175.

McGrath, P. T., Rockman, M. V., Zimmer, M., Jang, H., Macosko, E. Z., Kruglyak, L., & Bargmann, C. I. (2009). Quantitative mapping of a digenic

behavioral trait implicates globin variation in C. elegans sensory behaviors. *Neuron*, *61*(5), 692-699.

Mello, C. C., Kramer, J. M., Stinchcomb, D., & Ambros, V. (1991). Efficient gene transfer in C. elegans: extrachromosomal maintenance and integration of transforming sequences. *The EMBO journal*, *10*(12), 3959-3970.

Meng, L., Zhang, A., Jin, Y., & Yan, D. (2016). Regulation of neuronal axon specification by glia-neuron gap junctions in C. elegans. *Elife*, *5*, e19510.

Michael Hendrix, E., Myatt, L., Sellers, S., Russell, P. T., & Larsen, W. J. (1995). Steroid hormone regulation of rat myometrial gap junction formation: effects on cx43 levels and trafficking. *Biology of reproduction*, *52*(3), 547-560.

Mimoto, A., Fujii, M., Usami, M., Shimamura, M., Hirabayashi, N., Kaneko, T., Sasagawa, N., & Ishiura, S. (2007). Identification of an estrogenic hormone receptor in Caenorhabditis elegans. *Biochemical and biophysical research communications*, *364*(4), 883-888.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nature methods*, *19*(6), 679-682.

Oikonomou, G., & Shaham, S. (2011). The glia of Caenorhabditis elegans. *Glia*, *59*(9), 1253-1263.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, *25*(13), 1605-1612.

Pluciennik, F., Verrecchia, F., Bastide, B., Herve, J., Joffre, M., & Deleze, J. (1996). Reversible interruption of gap junctional communication by testosterone propionate in cultured Sertoli cells and cardiac myocytes. *The Journal of membrane biology*, *149*, 169-177.

Proctor, H., & Owens, I. I. (2000). Mites and birds: diversity, parasitism and coevolution. *Trends Ecol Evol*, *15*(9), 358-364.

Reed, E. M., & Wallace, H. R. (1965). Leaping Locomotion by an Insect-Parasitic Nematode. *Nature*, *206*(4980), 210-&.

Risek, B., Guthrie, S., Kumar, N., & Gilula, N. B. (1990). Modulation of gap junction transcript and protein expression during pregnancy in the rat. *The Journal of cell biology*, *110*(2), 269-282.

Saul-Gershenz, L. S., & Millar, J. G. (2006). Phoretic nest parasites use sexual deception to obtain transport to their host's nest. *Proc Natl Acad Sci U S A*, *103*(38), 14039-14044.

Schroeder, N. E., Androwski, R. J., Rashid, A., Lee, H., Lee, J., & Barr, M. M. (2013). Dauer-specific dendrite arborization in C. elegans is regulated by KPC-1/Furin. *Curr Biol*, *23*(16), 1527-1535.

Singhvi, A., & Shaham, S. (2019). Glia-neuron interactions in Caenorhabditis elegans. *Annual Review of Neuroscience*, *42*, 149-168.

Stern, D. L. (2014). Identification of loci that cause phenotypic variation in diverse species with the reciprocal hemizygosity test. *TRENDS in Genetics*, *30*(12), 547-554.

Stout Jr, R. F., Verkhratsky, A., & Parpura, V. (2014). Caenorhabditis elegans glia modulate neuronal activity and behavior. *Frontiers in cellular neuroscience*, *8*, 67.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585-595.

Taylor, S. R., Santpere, G., Weinreb, A., Barrett, A., Reilly, M. B., Xu, C. A., Varol, E., Oikonomou, P., Glenwinkel, L., McWhirter, R., Poff, A., Basavaraju, M., Rafi, I., Yemini, E., Cook, S. J., Abrams, A., Vidal, B., Cros, C., Tavazoie, S., . . . Miller, D. M. (2021). Molecular topography of an entire nervous system. *Cell*, *184*(16), 4329-+.

Tominaga, N., Ura, K., Kawakami, M., Kawaguchi, T., Kohra, S., Mitsui, Y., Iguchi, T., & Arizono, K. (2003). Caenorhabditis elegans responses to specific steroid hormones. *Journal of health science*, *49*(1), 28-33.

White, J. G., Southgate, E., Thomson, J. N., & Brenner, S. (1986). The structure of the nervous system of the nematode Caenorhabditis elegans. *Philos Trans R Soc Lond B Biol Sci*, *314*(1165), 1-340.

White, P. S., Morran, L., & de Roode, J. (2017). Phoresy. *Current biology*, *27*(12), R578-R580.

Widmayer, S. J., Evans, K. S., Zdraljevic, S., & Andersen, E. C. (2022). Evaluating the power and limitations of genome-wide association studies in Caenorhabditis elegans. *G3*, *12*(7), jkac114.

Yim, H., Choe, D. T., Bae, J. A., Kang, H.-M., Nguyen, K. C., Choi, M.-k., Ahn, S., Bahn, S.-k., Yang, H., & Hall, D. H. (2023). Mind of a dauer: Comparative connectomics reveals developmental plasticity. *bioRxiv*, 2023.2003. 2023.533915.

Yu, W., Dahl, G., & Werner, R. (1994). The connexin43 gene is responsive to oestrogen. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *255*(1343), 125-132.

Zhang, G., Roberto, N. M., Lee, D., Hahnel, S. R., & Andersen, E. C. (2022). The impact of species-wide gene expression variation on Caenorhabditis elegans complex traits. *Nat Commun*, *13*(1), 3462.

Zhang, M. G., & Sternberg, P. W. (2022). Both entry to and exit from diapause arrest in Caenorhabditis elegans are regulated by a steroid hormone pathway. *Development*, *149*(9), dev200173.

국문 초록

닉테이션 행동의 다양성에 대한 분자유전학적 연구

양희승

서울대학교 생명과학부

행동은 생물체에서 가장 복잡한 표현형 중 하나로, 유전자에서부터 환경에 이르기까지 수많은 요소가 행동 조절에 관여한다. 유전적인 차이가 어떻게 행동의 다양성에 관여할 수 있는지는 유전법칙과 유전물질이 발견된 이래로 생물학의 주요 질문 중 하나였다. 본 연구에서는 예쁜꼬마선충을 이용하여, 닉테이션 행동의 다양성의 근간이 되는 자연상의 유전변이를 발굴하고자 하였다. 예쁜꼬마선충 유충은 척박한 환경에서 자라면 생존에 유리한 여러 가지 특징을 가진 다우어라는 대체 발달단계로 접어든다. 다우어는 발생단계 특이적으로 닉테이션이라는 행동을 보이는데, 이는 다른 종과의 상호작용을 유도하며 다우어가 편승을 통해 다른 서식지로 분산할 수 있도록 촉진하는 행동이다. 본 연구에서는 전 세계에서 채집된 139 가지 야생형 예쁜꼬마선충의 닉테이션을 측정하였고, 그 비율에 차이가 있음을 확인했다. 전장유전체 연관분석을 포함한 여러가지 분석기법을 사용하여,

인간의 하이드록시 스테로이드 탈수소효소의 상동체인 *nta-1* 유전자의 변이가 야생형 간의 닉테이션 다양성을 조절한다는 것을 입증했다. 또한 형질 전환 동물을 활용하여 *nta-1* 의 프로모터 영역의 변이가 다양한 조직, 특히 중배엽에서 유래한 교세포인 GLR 세포에서의 발현을 조절하여 닉테이션 다양성을 유도하는 것을 확인할 수 있었다. 이러한 결과들은 유전자의 코딩 영역의 서열 변화가 아닌 프로모터 영역의 서열 변이가 어떻게 진화적 신기성에 기여할 수 있는가를 보여주는 예시를 제공하였다. 나아가서 *nta-1* 은 지역에 따라 다른 대립유전자 분포 양상을 보였으며, 집단유전학적 분석 결과는 *nta-1* 의 다양성이 균형 선택을 통해 유지되어 온 예쁜꼬마선충 집단의 고전적인 다양성에 해당함을 보였다. *nta-1* 은 닉테이션 뿐만 아니라 다우어를 거친 이후의 생식 속도에 영향을 미쳤으며, 이는 분산과 생식에 대한 *nta-1* 의 다면발현이 *nta-1* 대립유전자의 균형 선택에 영향을 미쳤음을 시사한다.

주요어 : 예쁜꼬마선충, 유전학, 닉테이션, 자연 변이, 행동 다양성, 전장유전체 연관분석

학 번 : 2015-20440