Review

CellPress

The laboratory domestication of *Caenorhabditis elegans*

Mark G. Sterken[1,2], L. Basten Snoek[1], Jan E. Kammenga[1], and Erik C. Andersen[3]

[1] Laboratory of Nematology, Wageningen University, Droevendaalsesteeg 1, 6708 PB, Wageningen, The Netherlands
[2] Laboratory of Virology, Wageningen University, Droevendaalsesteeg 1, 6708 PB, Wageningen, The Netherlands
[3] Department of Molecular Biosciences, Northwestern University, Evanston, IL, USA

Model organisms are of great importance to our understanding of basic biology and to making advances in biomedical research. However, the influence of laboratory cultivation on these organisms is underappreciated, and especially how that environment can affect research outcomes. Recent experiments led to insights into how the widely used laboratory reference strain of the nematode *Caenorhabditis elegans* compares with natural strains. Here we describe potential selective pressures that led to the fixation of laboratory-derived alleles for the genes *npr-1, glb-5,* and *nath-10*. These alleles influence a large number of traits, resulting in behaviors that affect experimental interpretations. Furthermore, strong phenotypic effects caused by these laboratory-derived alleles hinder the discovery of natural alleles. We highlight strategies to reduce the influence of laboratory-derived alleles and to harness the full power of *C. elegans*.

Model organisms pave the road to biological discovery

Sustained progress in the biological sciences is facilitated by discoveries using organisms amenable to laboratory investigation. They have large numbers of offspring, are small, and are easy to maintain. Many different species have these attributes. For example, the single-cell eukaryote *Saccharomyces cerevisiae* (baker's yeast) is immensely powerful as a genetic model organism for conserved cellular processes [1] and for quantitative genetics using large populations [2]. The fruit fly *Drosophila melanogaster* contributed extensively to our understanding of signal-transduction pathways and developmental patterning [3]. The free-living nematode *Caenorhabditis elegans* is a widely used model organism in studies of development [4,5], mechanistic neurobiology [6], aging [7], and small RNAs [8,9]. When the results from experimental studies of model organisms are tabulated, it is obvious that they facilitated much of what we know about conserved biological processes.

Quantitative geneticists often use tractable model organisms to identify loci and (sometimes) genetic variants that influence phenotypic differences among populations. To elucidate the underlying genetic basis of complex traits, recombinant offspring are generated and their traits measured. Organisms that give rise to large (preferably clonal) populations and are easy to manipulate experimentally enable these approaches. These attributes make *S. cerevisiae* the most powerful eukaryotic organism for quantitative genetics [2,10]. However, as a metazoan genetic system, *C. elegans* is unmatched [11]. It has an extremely rapid life cycle (3.5 days at 20°C), produces 200–300 offspring per hermaphrodite individual, possesses a small and well-annotated genome, and can be

Glossary

Axenic culture: conditions where organisms are grown in completely synthetic media. In the case of *Caenorhabditis elegans*, media are based on a liver extract [63,64].
Dauer: second larval stage (L2) animals enter this alternate larval stage at high culture density, low food abundance, and high temperature [Altun, Z.F. *et al.* (2002 Z.F. *WormAtlas* (http://www.wormatlas.org/)]. These L3 dauer larvae can survive stressful conditions and are thought to disperse to new locations in nature.
Ecological niche: the specific environment in which an organism lives and competes for resources. *C. elegans* is most often found in decaying material and not in soil [15,31].
FLP-18: one of two FMRFamide neuropeptides encoded by the *C. elegans* genome that can bind to the NPR-1 neuropeptide receptor. FLP-18 can activate the NPR-1(215V) allele (found in N2 animals) but not the NPR-1(215F allele) (found in all wild strains) [16,46].
FLP-21: the second of two FMRFamide neuropeptides encoded by the *C. elegans* genome that can bind to the NPR-1 neuropeptide receptor. FLP-21 is the natural ligand of NPR-1 [16,40].
Genome-wide association study (GWAS): a technique used on natural populations to identify genomic regions correlated with differences in phenotypic traits [12–14].
GLB-5: a globin domain protein that modifies behavioral responses to oxygen and oxygen/carbon dioxide stimuli [16,55].
Heritability: the amount of trait variation in a population that can be explained by genetic factors.
NATH-10: a vertebrate N-acetyltransferase homolog that has been shown to affect vulval induction in *C. elegans*. The 746I allele (N2) also results in faster maturation and greater fitness under laboratory conditions [35].
Neomorphic: describes a type of phenotype caused by an alteration of gene function that is novel and different from the normal function of the gene.
NPR-1: a G protein-coupled neuropeptide receptor normally activated by FLP-21. The 215V allele (N2) gained the ability to respond to FLP-18. The 215V allele affects many different traits [16,26,30,38,40,43–55].
Private alleles: specific alleles occurring in a single strain only.
Quantitative trait locus (QTL): a locus correlated with quantitative trait variation. For example, the *npr-1* QTL is correlated with variation in aggregation behavior.
Quantitative trait nucleotide (QTN): the variant site that causes variation in the quantitative trait.
RMG interneuron: the central neuron involved in most behaviors mediated by NPR-1 [47].
***Trans-band*:** in expression QTL studies, when variation of expression in many genes is correlated with the same genomic locus [30,59–61].
Vulval cell induction: when any of six hypodermal cells located on the ventral surface of the hermaphrodite are specified and divide to become vulval cells.

Corresponding authors: Kammenga, J.E. (jan.kammenga@wur.nl); Andersen, E.C. (erik.andersen@northwestern.edu).
Keywords: Caenorhabditis elegans; laboratory derived; selection; *npr-1*; aerotaxis.

0168-9525/
 http://dx.doi.org/10.1016/j.tig.2015.02.009

cryopreserved; in addition, transgenic strains are easily obtained. Wild strains isolated from nature can be phenotyped and genotyped to perform genome-wide association studies (GWASs) (see Glossary) [12–15]. This combination of studies on natural allelic variation paired with analyses of mutations using the laboratory strain offer a powerful approach to broaden our understanding of how genetic background contributes to phenotype. However, characterization of the behaviors and genomes of wild *C. elegans* strains led to suspicions about laboratory adaptation in the widely used N2 laboratory strain [16]. Wild type strains used in other model organisms have laboratory-derived variants that result in large pleiotropic effects, including cell clumping in *S. cerevisiae* [17,18] and plant growth in *Arabidopsis thaliana* [19].

Here we review documented examples of *C. elegans* laboratory-derived alleles in the commonly used Bristol (or N2) strain and the effects on its phenotype. We first describe how the laboratory strain N2 is different from wild strains of *C. elegans* and discuss the laboratory history of this nematode to gain insight into genetic bottlenecks and possible laboratory selection. Then we discuss three known laboratory-derived alleles and their effects on *C. elegans* biology, including implications for the interpretation of observations that can confound experimental outcomes.

N2 is distinct from all wild strains of *C. elegans*

Since its introduction to the research community by Sydney Brenner in 1974 [20], the Bristol (or N2) strain has been used in many laboratories and became the canonical wild type strain (for more history, see Box 1). From the Laboratory of Molecular Biology in Cambridge, N2 spread across the world via trainees from the Brenner laboratory, resulting in massive clonal amplification of N2 around the world. However, before its dissemination and cryogenic preservation, the N2 strain was propagated for many generations leading to the accumulation and selection of random mutations. We do not know how often the strain was transferred to new cultures during this early propagation. Conservatively, the strain could have been passaged every 2 months. At the other extreme, the strain could have been passaged every 4 days. Therefore, the strain underwent approximately 300 to 2000 generations from 1951 to 1969 (Box 1). Given the germline mutation rate of 2.7×10^{-9} mutations per site per generation [21], up to 1000 neutral mutations could have accumulated before cryogenic preservation. Furthermore, additional genetic differences among N2 strains from different laboratories arose after dispersal of this strain around the world, causing differences in the phenotypes of these standard wild type strains [22,23].

The cultivation history of the Bristol N2 strain provides only a few opportunities to identify mutations that accumulated during the early culturing period (Box 1 and Figure 1). Clues come from a strain that diverged from N2 some time before 1963 while in the Dougherty laboratory [Dougherty, E.C. (1963) Letter from Ellsworth C. Dougherty to Sydney Brenner. *CSHL Archives Repository* (http://libgallery.cshl.edu/items/show/60761)] up to 12 years after initial isolation from nature. This strain was mislabeled as *Caenorhabditis briggsae* – a mistake that was corrected later [24]. In 1995 and 2009, hermaphrodites were removed from axenic culture and frozen as the LSJ1 and LSJ2 strains [24,25], respectively. Comparison by sequencing revealed an estimate of approximately 100 accumulated variants in N2 [24]. However, no strains are currently known that diverged from N2 before the LSJ1 and LSJ2 strains diverged. Therefore, it is impossible to identify the mutations that accumulated in the initial decade after isolation.

From analysis of the genotypes and phenotypes of wild strains, we understand a great deal about variation in nature. Notably, many *C. elegans* strains reported to be isolated from nature were contaminated by the N2 strain (Table 1) [16]. Initial characterizations of natural phenotypic variation were confounded by these contaminated strains [26,27]. Fortunately, recent sampling and genotyping of true wild strains have made it possible to study natural variation in *C. elegans* [12,28,29]. The genomes of most wild *C. elegans* strains isolated from nature are highly related, sharing nearly two-thirds of the genome. This high degree of sharing is likely to be the effect of advantageous alleles that swept through the population reducing linked variation [12] and background selection that eliminates variation linked to deleterious alleles [30]. However, we still cannot identify all of the alleles that accumulated during laboratory propagation of the N2 strain, even with these genotype data. The N2 genome contains private alleles found in the strain when isolated from nature along with mutations that accumulated during laboratory propagation. Together, this mix of alleles

Box 1. *Caenorhabditis elegans*: the journey from nature to the bench

Most *C. elegans* research laboratories use the strain named N2, which was collected in 1951 from mushroom compost in Bristol, UK. Like most model organisms, the journey from nature to the laboratory was circuitous (see Figure 1 in main text). The compost was collected by L.N. Staniland, who brought the sample to a short course on plant nematology organized by the Ministry of Agriculture and Fisheries [16]. From this sample, Bristol *C. elegans* was isolated by Warwick Nicholas, who cultured the animal first on Petri dishes containing nutrient agar with bacterial contaminants as food [16]. Later, Nicholas developed axenic liquid cultures from the nutrient agar cultures, as these conditions required less frequent subculturing [64]. In 1957, the nematodes were transported to the laboratory of Ellsworth Dougherty at the Kaiser Foundation Research Institute in Richmond, CA in a liquid axenic culture [16]. In the Dougherty laboratory, Eder Hansen cultured the nematodes. Two types of culture were established: nutrient agar slants seeded with *Escherichia coli* in test tubes and liquid axenic culture based on liver extract [63].

Concurrently, Sydney Brenner sought an organism suitable for neurobiology research [65]. He corresponded with Dougherty and even isolated nematodes from his own garden [66]. This nematode culture was called the N1 strain. Brenner requested the Bristol strain from Dougherty and it was sent in 1963 [20,65] [Dougherty, E.C. (1963) Letter from Ellsworth C. Dougherty to Sydney Brenner. *CSHL Archives Repository* (http://libgallery.cshl.edu/items/show/60761)]. In the Brenner laboratory, the liquid axenic culture was transferred to agar plates containing *E. coli*. After several passages of a population containing both males and hermaphrodites, a single hermaphrodite was selected. This strain, which was used for all subsequent work, was called N2 [66]. The populations were kept in culture on *E. coli* monoxenic agar plates and the hermaphrodite strain was eventually frozen in 1969 by John Sulston [67].

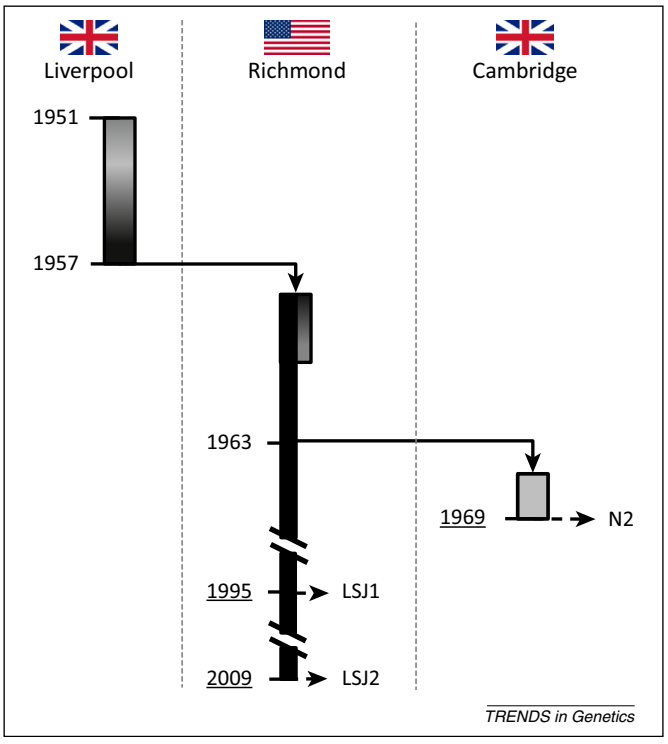

Figure 1. The history of the Bristol (N2) lineage. Monoxenic (gray) and axenic (black) cultures are denoted by colored boxes. The gradients indicate uncertainty in when the culture type was switched. The dates show the year of isolation or when the strain was moved to another laboratory. Underscores mark the dates of cryogenic preservation. In 1951, Bristol was isolated by L.N. Staniland and Warwick Nicholas. The strain was kept in Liverpool first as a monoxenic culture and later as an axenic culture [16]. In 1957, it was shipped to the Kaiser Foundation Research Institute in Richmond, CA. During early laboratory propagation, both axenic and monoxenic cultures were maintained. Later, the monoxenic culture was discontinued and the axenic culture continued [Dougherty, E.C. (1963) Letter from Ellsworth C. Dougherty to Sydney Brenner. *CSHL Archives Repository* (http://libgallery.cshl.edu/items/show/60761)]. The LSJ1/LSJ2 strains originate from this axenic culture [16,25], as does N2. It is unclear when exactly the LSJ1/LSJ2 lineage split from N2 [24]. In 1963, Sydney Brenner received an axenic culture containing *Caenorhabditis elegans* [Dougherty, E.C. (1963) Letter from Ellsworth C. Dougherty to Sydney Brenner. *CSHL Archives Repository* (http://libgallery.cshl.edu/items/show/60761)], which was cryogenically preserved by John Sulston in 1969 [67]. The LSJ1 strain was cryogenically preserved in 1995 [16,25] and the LSJ2 strain in 2009 [16].

makes it impossible to identify laboratory-derived alleles from sequence information alone.

Selective pressures: nature versus laboratory

To understand how laboratory conditions could influence *C. elegans*, we need to know more about its natural habitat and ecology. Although progress has been made in recent years, the ecology of *C. elegans* remains largely unknown [15,31,32]. Despite frequent use of the statement in the literature, it is unlikely that *C. elegans* is a soil nematode. Soil samples harbor *C. elegans* only when in close proximity to rotting vegetation or fruit [15,33] and recent successful sampling suggests that its natural habitat is rotting material. Wild strains were isolated successfully from rotting hogweed [15,28], rotting fruits [15,28,31,33], and compost [28,31]. Additionally, strains have been isolated from 'carrier' species such as snails and terrestrial isopods [15]. Current observations indicate that *C. elegans* occupies short-lived, microbiota-rich habitats. In this niche, it establishes a population quickly and is thought to compete for bacterial food with other species [15,31]. When food is limiting and population density is high, *C. elegans* enters a long-lived alternative larval stage called dauer. These dauers are likely to endure periods without food while dispersing to new habitats [34]. By contrast, laboratory cultivation provides a much more constant environment (Box 2).

When animals are removed from their natural environments and transported to the laboratory, species undergo strong selective pressures that ultimately can change the organism. The impact of a laboratory environment on an

Table 1. A large number of 'wild' *Caenorhabditis elegans* strains are mislabeled N2 strains or recombinant strains derived from N2

Strain	Genotype[a]	Refs
CB3191	N2	[16,68]
CB3192	N2	[16,68]
CB3193	N2	[16,68]
CB3194	N2	[16,68]
CB3195	N2	[16,68]
CB4507	N2	[16]
CB4555	N2×CB4851 recombinant	[16]
DH424	N2×CB4851 recombinant	[16]
DR1349	N2×CB4851 recombinant	[16]
PX176	N2	[16]
TR388	N2	[16,68]
TR389	N2	[16,68]
TR403	N2×CB4851 recombinant	[16]

[a]All strains with the N2 genotype have N2 markers at 1453 of 1454 markers spread throughout the genome. The N2×CB4851 recombinant strains are largely or completely N2 for chromosomes I, II, III, and X and CB4851 for chromosomes IV and V. Strain TR389 has N2 for 1453 markers but harbors the CB4856-like *glb-5* deletion allele.

Box 2. Living conditions of *Caenorhabditis elegans* in the laboratory

The life cycle of *C. elegans* comprises an embryonic stage followed by four larval stages (L1–L4) and an adult stage. The N2 strain completes one generation every 3.5 days at 20°C. Alternatively, *C. elegans* can enter a long-term survival stage (dauer) as an alternative to the standard L3 larval stage [Altun, Z.F. *et al.* (2002–2014) *WormAtlas* (http://www.wormatlas.org/)].

Axenic culture

Axenic cultures do not contain other organisms as a food source and can be chemically defined or contain extracts of organic material (e.g., liver). Such cultures can be either in a solid state (e.g., nutrient agar) or in liquid.

Axenic cultures are now not often used for keeping *C. elegans*, with the exception of transport into space [69]. In the early days of *Caenorhabditis* sp. research, much time was invested in establishing a defined axenic medium to grow nematodes [63,64], for two major reasons. First, it required a lower frequency of subculturing. Before the cryopreservation method was developed, infrequent subculture requirements were a great advantage. Second, axenic culture offered the ability to chemically define the medium, which allows the researcher to alter components and investigate nutritional requirements.

Monoxenic culture

Monoxenic cultures contain one organism as a food source. In the case of *C. elegans*, the nematode is almost exclusively cultured on media containing *Escherichia coli*.

There are two main methods for monoxenic culture of *C. elegans*: in liquid or on solid medium. In liquid culture, animals are grown with agitation in solution. On solid media, the animals are kept on nematode growth medium (NGM) agar plates seeded with an *E. coli* strain [20].

organism is significant: environmental conditions are kept nearly constant, breeding regimens are strictly enforced, and food is readily available (Box 2). Additionally, researchers impose novel pressures by the culturing system; for example, transferring individual animals to start a new culture (bottlenecks). The substrate on which animals are grown should be considered. Agar plates offer a 2D substrate whereas rotting fruit is a 3D environment [15]. This laboratory propagation results in evolution through artificial selection, which inevitably affects genotypic and phenotypic characteristics of model organisms, including *C. elegans*.

From two studies, it is clear that the N2 genotype exhibits higher fitness in laboratory conditions than wild strains [25,35]. The phenotype of N2 is distinct from wild strains in several ways, including aggregation behavior, maturation time, fecundity, body size, and many other traits [25,26,28,35–42]. The atmospheric oxygen concentration on agar plates is substantially higher than levels preferred by wild strains [43–45] and laboratory oxygen concentration is a strong selective pressure on the organism. This oxygen concentration affects the growth and physiology of the animal profoundly, because many behaviors are altered, including how the animals consume bacterial food. These oxygen-dependent effects are so profound that two of three confirmed laboratory-derived alleles are associated with altered behaviors at higher oxygen concentrations [16,40]. The effects of these alleles, and possibly other laboratory-derived alleles, are pleiotropic, so they could have been selected by additional, unexplained pressures.

Laboratory-derived alleles in the *C. elegans* N2 strain and their functional consequences

During the first 18 years that the N2 strain was grown in the laboratory, many mutations arose that might not have conferred any selective advantage [21]. However, we know that laboratory propagation of this strain led to the fixation of several alleles that confer a strong selective advantage under these conditions [25,35]. Laboratory-derived alleles are random mutations that increase the fitness of the organism under laboratory conditions. At least three genes in the N2 strain have laboratory-derived variation: *npr-1*, *glb-5*, and *nath-10* [16,35,40]. For each of these genes, the N2 genome contains a variant that differs from that found among all *bona fide* wild strains. Furthermore, the two N2-diverged strains, LSJ1 and LSJ2 (Figure 1), carry the same alleles as wild strains. These results provide further evidence for the laboratory origin of the alleles, because LSJ1 and LSJ2 were separated from the N2 strain at least 6 years before cryopreservation [16,24,25].

The neuropeptide receptor encoding gene npr-1: laboratory adaptation abnormally represses the C. elegans nervous system

A seven-transmembrane neuropeptide receptor encoded by *npr-1* was first identified as a master regulator of a behavioral dimorphism where animals either aggregate or remain solitary in the presence of bacterial food [26]. This aggregation behavior mapped to an amino acid substitution within the third intracellular loop of the NPR-1 receptor. Wild strains of *C. elegans* contain the 215F allele, with which the NPR-1 receptor responds to the neuropeptide FLP-21. By contrast, the laboratory strain N2 contains the 215V allele, which leads to a neomorphic gain-of-function sensitivity of NPR-1 to FLP-18 in addition to sensitivity to FLP-21 [46]. This gain-of-function sensitivity creates an abnormally repressed neural circuit through inactivation of the RMG interneuron [47], affecting a large number of behaviors (Table 2) [16,26,30,38,40,43–55].

Table 2. The laboratory-derived allele of *npr-1* causes a large number of phenotypic effects

Trait	Phenotypic effect	Related to aerotaxis?	Refs
Aggregation	Lower	Yes	[26,43,46,47]
Taxis to low oxygen	Lower	Yes	[40,43–45]
Pathogen avoidance	Higher	Yes	[40,48,49]
Lifetime fecundity	Higher	Yes	[40]
Body size	Larger	Yes	[40]
Gene-expression regulation	NA	Yes	[30,40]
Ethanol tolerance	Lower	Not tested	[50]
Carbon dioxide avoidance	Higher	Not tested	[16,51]
Heat avoidance	Higher	Not tested	[52]
Hermaphrodite leaving	Lower	Yes	[38]
Pheromone responses	Repulsed	Not tested	[47,53]
Lethargus quiescence	Higher	Yes	[54]
Crawling speed	Lower	Not tested	[55]

A modified aerotaxis response is one of the central drivers of the behavioral differences caused by variation in NPR-1 (Figure 2). Wild type *C. elegans* strongly prefer oxygen concentrations lower than ambient levels [43–45]. On agar plate cultures, this behavior manifests as taxis to oxygen concentrations of approximately 10%, which are often found at the border of the bacterial lawn [43–45]. Aggregation of animals decreases the local oxygen concentration even further [45]. This reduction in oxygen concentration caused by aggregation reinforces the further formation of

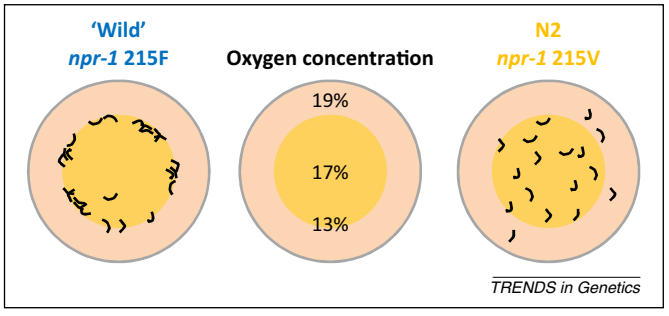

Figure 2. The aerotaxis effects of *npr-1*. Natural *Caenorhabditis elegans* strains aggregate at the edges of the bacterial lawn (orange) when propagated in laboratory conditions on monoxenic agar plates (left). The edges of the bacterial lawn have lower-than-ambient oxygen concentrations (approximately 13%, center). Wild *C. elegans* strains respond to this oxygen gradient and prefer lower oxygen concentrations in the presence of bacterial food [43]. The abnormal N2 strain is less sensitive to oxygen concentration and does not aggregate at the edges of the bacterial lawn (right). This difference in aerotaxis, or oxygen preference, leads to different aggregation and lawn-leaving behaviors [26,38,43,46,47]. Because of these behavioral changes, strains also differ in exposure to pathogens [40,48,49]. Additionally, the aggregation behavior might cause a chronic mild starvation state, which results in reduced growth rate [40], reduced fecundity [40], altered gene expression [30,40], increased crawling speed [55], and reduced quiescence during lethargus [54].

aggregates, which in turn decreases available food as animals compete in close proximity. The reduction in growth rate and offspring production observed in wild *C. elegans* strains is likely to be caused by a mild starvation state in aggregates [40]. Additionally, these animals could experience higher levels of pheromones, potentially signaling a stress state that reduces growth rate and offspring production [40]. The attraction of wild strains to the border of the lawn increases the exposure to bacteria [38]. When these bacteria are pathogenic, strains with the 215F allele will be exposed more extensively to the pathogen and succumb faster to infection than the N2 strain with the 215V allele [40,48,49].

Most traits related to *npr-1* variation are linked to aerotaxis behavior (Table 2). Traits not linked to aerotaxis include heat avoidance [52], ethanol tolerance [50], carbon dioxide avoidance [16,51], and pheromone response [47,53]. However, these traits still might be linked to aerotaxis via the RMG neuron, but these connections have not been characterized extensively. For example, the ethanol response might be regulated through FLP-18 [50] or might be sensed in the nociception neuron ASH, which is connected to RMG [47]. RMG is also connected to the pheromone-sensing ADL neuron. As the aggregation observed in wild strains is likely to cause higher pheromone exposure and lower oxygen concentration, it is difficult to distinguish the contributions of the two factors [40]. Many traits where *npr-1* variation is implicated in the behavior have not been directly connected to aerotaxis behaviors by empirical evidence. Most of these traits, however, are likely to be caused by variation in the aerotaxis responses and differences in food consumption mediated by RMG through its role as a 'hub-and-spoke' neuron [47].

Other laboratory-derived alleles found in the N2 strain affect nath-10 and glb-5

Together with NPR-1, the neuronal globulin domain protein GLB-5 affects a behavioral response to changes in carbon dioxide and oxygen concentrations. The causal variant in *glb-5* is a duplication/insertion of 765 base pairs leading to a 179-amino acid truncation and a 40-amino acid substitution in the N2 strain [16,55]. The combination of laboratory-adapted alleles at the *glb-5* and *npr-1* loci leads to opposing responses to changes in carbon dioxide concentration compared with the wild alleles. Wild strains move more quickly and make more turns when they sense a simultaneous decrease in carbon dioxide concentration and increase in oxygen concentration. By contrast, N2 animals move more quickly and make more turns when they sense an increase in carbon dioxide concentration. Furthermore, the N2 allele of *glb-5* desensitizes the URX neuron (which is also connected to RMG) to small fluctuations in oxygen levels, leading to reduced responses to oxygen concentration [16,55]. The *npr-1* and *glb-5* alleles exhibit a genetic interaction. A strain with the natural alleles at both loci displays a phenotype that differs from that of the strains with only one natural allele. If only the N2 allele of *npr-1* is present, animals will react to oxygen in a concentration-dependent manner, whereas the N2 *glb-5* allele on its own renders them insensitive to fluctuations in oxygen concentration. If animals carry both alleles, however, they react strongly to minor shifts in oxygen and carbon dioxide concentrations around the atmospheric oxygen concentration [16,55]. These discoveries related to oxygen and carbon dioxide preferences led to original observations of the derived nature of the N2 strain [16].

Variation in the human N-acetyltransferase homolog gene (*nath-10*) causes variation in vulval cell-fate specification and shows pleiotropic effects on fecundity and growth rate [35]. The laboratory-derived allele encodes a putative substitution of methionine-746 with isoleucine in a highly conserved region of the N-acetyltransferase domain. This laboratory-derived allele was identified because of specific effects on variation in vulval cell-fate specification. Variation in *nath-10* causes visible effects on vulval development only when additional mutations sensitize *let-60* Ras pathway activity. The laboratory-derived allele of *nath-10* partially suppresses a lower level of vulval cell-fate induction caused by a reduction-of-function mutation in the gene encoding an epidermal growth factor (EGF) receptor (*let-23*) and enhances the level of vulval cell-fate induction caused by a gain-of-function mutation in the gene encoding Ras (*let-60*), indicating that the laboratory-derived allele of *nath-10* stimulates Ras pathway activity. This allele also affects age at maturity, brood size, and egg-laying speed through an increase in the production of sperm. Given this large effect on fitness, the N2 allele of *nath-10* causes a selective advantage when animals are grown in laboratory competition assays [35].

The effects of natural allelic variation is obscured by propagation of strains in the laboratory

To investigate the effects of laboratory alleles, we analyzed the *C. elegans* linkage mapping results from the past decade for linkage to *npr-1*, *glb-5*, and *nath-10* genomic regions (Table 3). A large number of linkage mapping experiments detected a quantitative trait locus (QTL) with a confidence interval that includes the *npr-1* locus, including dauer formation [56,57], body size [37,40], lifespan [58], and vulval index [35]. Laboratory-derived alleles have large effects when strains are grown in laboratory conditions. To estimate this effect, we compared the broad-sense heritability (H^2) and the variance explained by the *npr-1* QTL. This comparison indicates how much of the genetic difference among strains is influenced by the *npr-1* QTL. Variation at the *npr-1* locus explains 30–82% of the variance contributed by genetic factors for various traits [37,40] – a large phenotypic effect. However, not all traits consistently detect a QTL at the *npr-1* locus. For example, one expression QTL study detected a *trans*-band at *npr-1* [30] but three other studies did not [59–61]. Similarly, one study detected the *npr-1* QTL for fecundity [40] but two other studies did not [36,37]. We suggest that studies that failed to detect a QTL near *npr-1* are likely to have had different laboratory culture conditions, including the number of animals in the culture and the assay temperature. The differences in population density and variation at *npr-1* interact with large effects on the growth and physiology of the organism [40]. With increased culture density comes more crowding of animals at the edge of the bacterial lawn. This crowding causes a chronic

Table 3. Many linkage mapping genetic studies identified the *npr-1*, *glb-5*, or *nath-10* locus

Strains	Trait	Interval detected*[a]			Identified causal gene	Refs
		nath-10	*glb-5*	*npr-1*		
N2×BO	Lifespan		Yes			[70]
	Oxidative stress response		Yes			
N2×CB4856	Age at maturity, 24°C	Yes				[36]
N2×CB4856	Body mass, 12°C			Yes		[37]
	Body mass, 24°C			Yes		
N2×DR1350	Dauer formation, high food, 19°C			Yes		[56]
	Dauer formation, food, plasticity			Yes		
N2×CB4856	Pathogen susceptibility			Yes	*npr-1*	[49]
N2×CB4856	Lifespan			Yes		[58]
N2×CB4856	Carbon dioxide up shift, oxygen down shift		Yes	Yes	*glb-5, npr-1*	[16]
	Carbon dioxide down shift, oxygen up shift		Yes	Yes	*glb-5, npr-1*	
N2×CB4856	Oxygen sensing and response		Yes	Yes	*glb-5, npr-1*	[55]
N2[b]×CB4856	Male tail phenotype, 13°C	Yes				[71]
N2×CB4856	Gene expression, L4, 24°C	Yes				[60]
	Gene expression, L4 and reproductive, 24°C	Yes				
N2×CB4856	Gene expression, young adult, 20°C			Yes		[30]
N2×CB4856	Population growth on RNAi (8/11 genes)	Yes			*ppw-1*	[72]
N2×CB4856	Lawn leaving			Yes	*tyra-3*	[38]
N2×CB4856	Heat avoidance			Yes	*npr-1*	[52]
JU605×JU606	Vulval induction, 20°C			Yes		[35]
	Vulval induction, 25.5°C	Yes			*nath-10*	
	Vulval induction, plasticity	Yes			*nath-10*	
N2×CB4856	Bordering		Yes		*exp-1*	[39]
N2×CB4856	Thermal preference			Yes		[73]
	Isothermal dispersion			Yes		
N2×CB4856	Dauer formation	Yes		Yes		[57]
N2×CB4856	Gene expression			Yes		[74]
N2×CB4856	Lifetime fecundity			Yes	*npr-1*	[40]
	Adult body size		Yes	Yes	*npr-1*	
	Susceptibility to *Staphylococcus aureus*			Yes	*npr-1*	
	Gene expression			Yes	*npr-1*	
N2×CB4856	Dauer formation (pheromone exposure)	Yes	Yes			[57]
	Dauer formation (food exhaustion)	Yes			*nath-10*	
N2×CB4856	Embryonic development	Yes				[75]

[a]When these intervals are not detected, this negative result could be caused by several factors. For example, this interval is not involved in the trait. Alternatively, the lack of detection could be caused by technical reasons, including presence of markers or statistical power.
[b]The strain used for constructing the recombinant inbred population was CB5362, a strain containing the *tra-2(ar221)* and the *xol-1(y9)* mutations in an N2 background.

low-level starvation state, which has large phenotypic effects. Additionally, differences in rearing temperature result in altered growth rates and population densities with similar effects.

The *nath-10* locus has been associated with several traits, including age at maturity [35,36], an expression QTL *trans*-band [60], and dauer formation [57,62]. It is difficult to assess the effect of the locus in general, as only one study reports the contribution to heritable variation (52%) [36]. The *trans*-band associated with the *nath-10* locus was measured in L4 and reproductive animals at 24°C. Therefore, it is likely that the developmental differences caused by *nath-10* result in gene-expression differences. Because *nath-10* is a pleiotropic locus implicated in fecundity and growth rate, these correlated QTLs are unsurprising.

In summary, the contribution of laboratory-derived alleles to heritable variation is large (30–82%) and seems to be environment dependent. It is important to consider the context in which traits are measured. Given that both *npr-1* and *glb-5* affect behavior at atmospheric oxygen concentration [16,40,43], many behavioral studies using the N2 strain under standard laboratory conditions might be difficult to interpret with respect to a normal behavioral circuit and natural behaviors.

Where do we go from here?

C. elegans is an essential model organism used to understand human biology. However, we need to be aware of the large and pleiotropic phenotypic effects caused by laboratory-derived alleles, especially those alleles present in the reference strain N2. These alleles can influence our conclusions and could alter the interpretation of results for the understanding of human biology, as they alter the natural physiology of *C. elegans*. However, investigators should not abandon the N2 strain. The large experimental toolkit and the decades of results obtained by the study of this one strain are invaluable. These advantages need to be tempered with the knowledge that the N2 strain has been bred in a single environment for a long time before cryopreservation. Laboratories whose research focuses exclusively on the N2 strain and mutant derivatives should consider

expanding to more natural *C. elegans* strains, especially when the focus includes traits that are influenced by population density (e.g., metabolism).

Newly isolated *C. elegans* that are cryopreserved as soon as possible after arrival in the laboratory are an untapped resource of genetic variants to expand the experimental power of *C. elegans* and its applicability to humans. For example, these strains can be added to the panel currently used for GWASs [12–14]. Additionally, new recombinant inbred line collections can be constructed using natural strains, which will greatly benefit quantitative genetic studies. One of the major strengths of *C. elegans* is the combination of wild strains and the accumulated knowledge from the study of the laboratory strain N2. This combination allows rapid screening of causal genes to understand evolutionary and ecological genetics as well as making a larger impact on biomedical science. The *C. elegans* research community is ready for the next round of rapid and important progress once natural strains are integrated into the existing genetic toolkit.

Acknowledgments

The authors thank Rachel Ankeny, Bob Horvitz, Patrick McGrath, John Sulston, and Diana Wall for discussions, further documentation, and additional information about the history of the *C. elegans* N2 strain. They are grateful to Roel Bevers, Daniel Cook, Patrick McGrath, Anne Morbach, Lisa van Sluijs, Robyn Tanny, and Stefan Zdraljevic for editorial comments on the manuscript. M.G.S. was funded by the Graduate School for Production Ecology and Resource Conservation (PE&RC). L.B.S. was funded by ALW grant 823.01.001. J.E.K. received support from the EU-funded project PANACEA (222936). E.C.A. received support from the Pew Charitable Trusts and National Institutes of Health (NIH) grant R01GM107227.

References

1 Dunham, M.J. and Fowler, D.M. (2013) Contemporary, yeast-based approaches to understanding human genetic variation. *Curr. Opin. Genet. Dev.* 23, 658–664
2 Bloom, J.S. *et al.* (2013) Finding the sources of missing heritability in a yeast cross. *Nature* 494, 234–237
3 Ejsmont, R.K. and Hassan, B.A. (2014) The little fly that could: wizardry and artistry of *Drosophila* genomics. *Genes* 5, 385–414
4 Jackson, B.M. and Eisenmann, D.M. (2012) β-Catenin-dependent Wnt signaling in *C. elegans*: teaching an old dog a new trick. *Cold Spring Harb. Perspect. Biol.* 4, a007948
5 Felix, M.A. and Barkoulas, M. (2012) Robustness and flexibility in nematode vulva development. *Trends Genet.* 28, 185–195
6 Husson, S.J. *et al.* (2013) Optogenetic manipulation of neural activity in *C. elegans*: from synapse to circuits and behaviour. *Biol. Cell* 105, 235–250
7 Kenyon, C. (2011) The first long-lived mutants: discovery of the insulin/IGF-1 pathway for ageing. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 366, 9–16
8 Fire, A. *et al.* (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806–811
9 Kaufman, E.J. and Miska, E.A. (2010) The microRNAs of *Caenorhabditis elegans*. *Semin. Cell Dev. Biol.* 21, 728–737
10 Ehrenreich, I.M. *et al.* (2009) Genetic dissection of complex traits in yeast: insights from studies of gene expression and other phenotypes in the BY×RM cross. *Cold Spring Harb. Symp. Quant. Biol.* 74, 145–153
11 Gaertner, B.E. and Phillips, P.C. (2010) *Caenorhabditis elegans* as a platform for molecular quantitative genetics and the systems biology of natural variation. *Genet. Res.* 92, 331–348
12 Andersen, E.C. *et al.* (2012) Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. *Nat. Genet.* 44, 285–290
13 Rockman, M.V. and Kruglyak, L. (2009) Recombinational landscape and population genomics of *Caenorhabditis elegans*. *PLoS Genet.* 5, e1000419
14 Ashe, A. *et al.* (2013) A deletion polymorphism in the *Caenorhabditis elegans* RIG-I homolog disables viral RNA dicing and antiviral immunity. *Elife* 2, e00994
15 Felix, M.A. and Braendle, C. (2010) The natural history of *Caenorhabditis elegans*. *Curr. Biol.* 20, R965–R969
16 McGrath, P.T. *et al.* (2009) Quantitative mapping of a digenic behavioral trait implicates globin variation in *C. elegans* sensory behaviors. *Neuron* 61, 692–699
17 Yvert, G. *et al.* (2003) *Trans*-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat. Genet.* 35, 57–64
18 Brem, R.B. *et al.* (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science* 296, 752–755
19 van Zanten, M. *et al.* (2009) The many functions of ERECTA. *Trends Plant Sci.* 14, 214–218
20 Brenner, S. (1974) The genetics of *Caenorhabditis elegans*. *Genetics* 77, 71–94
21 Denver, D.R. *et al.* (2009) A genome-wide view of *Caenorhabditis elegans* base-substitution mutation processes. *Proc. Natl. Acad. Sci. U.S.A.* 106, 16310–16314
22 Vergara, I.A. *et al.* (2009) Polymorphic segmental duplication in the nematode *Caenorhabditis elegans*. *BMC Genomics* 10, 329
23 Gems, D. and Riddle, D.L. (2000) Defining wild-type life span in *Caenorhabditis elegans*. *J. Gerontol. A: Biol. Sci. Med. Sci.* 55, B215–B219
24 McGrath, P.T. *et al.* (2011) Parallel evolution of domesticated *Caenorhabditis* species targets pheromone receptor genes. *Nature* 477, 321–325
25 Weber, K.P. *et al.* (2010) Whole genome sequencing highlights genetic changes associated with laboratory domestication of *C. elegans*. *PLoS ONE* 5, e13922
26 de Bono, M. and Bargmann, C.I. (1998) Natural variation in a neuropeptide Y receptor homolog modifies social behavior and food response in *C. elegans*. *Cell* 94, 679–689
27 Reiner, D.J. *et al.* (2008) *C. elegans* anaplastic lymphoma kinase ortholog SCD-2 controls dauer formation by modulating TGF-β signaling. *Curr. Biol.* 18, 1101–1109
28 Volkers, R.J. *et al.* (2013) Gene–environment and protein-degradation signatures characterize genomic and phenotypic diversity in wild *Caenorhabditis elegans* populations. *BMC Biol.* 11, 93
29 Thompson, O. *et al.* (2013) The million mutation project: a new approach to genetics in *Caenorhabditis elegans*. *Genome Res.* 23, 1749–1762
30 Rockman, M.V. *et al.* (2010) Selection at linked sites shapes heritable phenotypic variation in *C. elegans*. *Science* 330, 372–376
31 Petersen, C. *et al.* (2014) The prevalence of *Caenorhabditis elegans* across 1.5 years in selected North German locations: the importance of substrate type, abiotic parameters, and *Caenorhabditis* competitors. *BMC Ecol.* 14, 4
32 Petersen, C. *et al.* (2015) Why we need more ecology for genetic models such as *C. elegans*. *Trends Genet.* 31, 120–127
33 Kiontke, K.C. *et al.* (2011) A phylogeny and molecular barcodes for *Caenorhabditis*, with numerous new species from rotting fruits. *BMC Evol. Biol.* 11, 339
34 Lee, H. *et al.* (2012) Nictation, a dispersal behavior of the nematode *Caenorhabditis elegans*, is regulated by IL2 neurons. *Nat. Neurosci.* 15, 107–112
35 Duveau, F. and Felix, M.A. (2012) Role of pleiotropy in the evolution of a cryptic developmental variation in *Caenorhabditis elegans*. *PLoS Biol.* 10, e1001230
36 Gutteling, E.W. *et al.* (2007) Mapping phenotypic plasticity and genotype–environment interactions affecting life-history traits in *Caenorhabditis elegans*. *Heredity* 98, 28–37
37 Gutteling, E.W. *et al.* (2007) Environmental influence on the genetic correlations between life-history traits in *Caenorhabditis elegans*. *Heredity* 98, 206–213
38 Bendesky, A. *et al.* (2011) Catecholamine receptor polymorphisms affect decision-making in *C. elegans*. *Nature* 472, 313–318
39 Bendesky, A. *et al.* (2012) Long-range regulatory polymorphisms affecting a GABA receptor constitute a quantitative trait locus (QTL) for social behavior in *Caenorhabditis elegans*. *PLoS Genet.* 8, e1003157
40 Andersen, E.C. *et al.* (2014) A variant in the neuropeptide receptor npr-1 is a major determinant of *Caenorhabditis elegans* growth and physiology. *PLoS Genet.* 10, e1004156

41 Kammenga, J.E. *et al.* (2007) A *Caenorhabditis elegans* wild type defies the temperature–size rule owing to a single nucleotide polymorphism in tra-3. *PLoS Genet.* 3, e34
42 Snoek, L.B. *et al.* (2014) A rapid and massive gene expression shift marking adolescent transition in *C. elegans*. *Sci. Rep.* 4, 3912
43 Gray, J.M. *et al.* (2004) Oxygen sensation and social feeding mediated by a *C. elegans* guanylate cyclase homologue. *Nature* 430, 317–322
44 Chang, A.J. *et al.* (2006) A distributed chemosensory circuit for oxygen preference in *C. elegans*. *PLoS Biol.* 4, e274
45 Rogers, C. *et al.* (2006) Behavioral motifs and neural pathways coordinating O_2 responses and aggregation in *C. elegans*. *Curr. Biol.* 16, 649–659
46 Rogers, C. *et al.* (2003) Inhibition of *Caenorhabditis elegans* social feeding by FMRFamide-related peptide activation of NPR-1. *Nat. Neurosci.* 6, 1178–1185
47 Macosko, E.Z. *et al.* (2009) A hub-and-spoke circuit drives pheromone attraction and social behaviour in *C. elegans*. *Nature* 458, 1171–1175
48 Styer, K.L. *et al.* (2008) Innate immunity in *Caenorhabditis elegans* is regulated by neurons expressing NPR-1/GPCR. *Science* 322, 460–464
49 Reddy, K.C. *et al.* (2009) A polymorphism in npr-1 is a behavioral determinant of pathogen susceptibility in *C. elegans*. *Science* 323, 382–384
50 Davies, A.G. *et al.* (2004) Natural variation in the npr-1 gene modifies ethanol responses of wild strains of *C. elegans*. *Neuron* 42, 731–743
51 Hallem, E.A. and Sternberg, P.W. (2008) Acute carbon dioxide avoidance in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U.S.A.* 105, 8038–8043
52 Glauser, D.A. *et al.* (2011) Heat avoidance is regulated by transient receptor potential (TRP) channels and a neuropeptide signaling pathway in *Caenorhabditis elegans*. *Genetics* 188, 91–103
53 Jang, H. *et al.* (2012) Neuromodulatory state and sex specify alternative behaviors through antagonistic synaptic pathways in *C. elegans*. *Neuron* 75, 585–592
54 Choi, S. *et al.* (2013) Analysis of NPR-1 reveals a circuit mechanism for behavioral quiescence in *C. elegans*. *Neuron* 78, 869–880
55 Persson, A. *et al.* (2009) Natural variation in a neural globin tunes oxygen sensing in wild *Caenorhabditis elegans*. *Nature* 458, 1030–1033
56 Harvey, S.C. *et al.* (2008) Quantitative genetic analysis of life-history traits of *Caenorhabditis elegans* in stressful environments. *BMC Evol. Biol.* 8, 15
57 Green, J.W. *et al.* (2013) Genetic mapping of variation in dauer larvae development in growing populations of *Caenorhabditis elegans*. *Heredity* 111, 306–313
58 Doroszuk, A. *et al.* (2009) A genome-wide library of CB4856/N2 introgression lines of *Caenorhabditis elegans*. *Nucleic Acids Res.* 37, e110
59 Li, Y. *et al.* (2006) Mapping determinants of gene expression plasticity by genetical genomics in *C. elegans*. *PLoS Genet.* 2, e222
60 Vinuela, A. *et al.* (2010) Genome-wide gene expression regulation as a function of genotype and age in *C. elegans*. *Genome Res.* 20, 929–937
61 Li, Y. *et al.* (2010) Global genetic robustness of the alternative splicing machinery in *Caenorhabditis elegans*. *Genetics* 186, 405–410
62 Green, J.W. *et al.* (2014) Highly polygenic variation in environmental perception determines dauer larvae formation in growing populations of *Caenorhabditis elegans*. *PLoS ONE* 9, e112830
63 Hansen, E.L.Y. *et al.* (1960) Differential nutritional requirements for reproduction of two strains of *Caenorhabditis elegans* in axenic culture. *Nematologica* 5, 27–31
64 Nicholas, W.L. and McEntegart, E.M. (1957) A technique for obtaining axenic cultures of rhabditid nematodes. *J. Helminthol.* 31, 135–144
65 Ankeny, R.A. (2001) The natural history of *Caenorhabditis elegans* research. *Nat. Rev. Genet.* 2, 474–479
66 Brown, A. (2003) *In the Beginning was the Worm: Finding the Secrets of Life in a Tiny Hermaphrodite*, Columbia University Press
67 Riddle, D.L. (1997) *C. elegans II*, Cold Spring Harbor Laboratory Press
68 Haber, M. *et al.* (2005) Evolutionary history of *Caenorhabditis elegans* inferred from microsatellites: evidence for spatial and temporal genetic differentiation and the occurrence of outbreeding. *Mol. Biol. Evol.* 22, 160–173
69 Adenle, A.A. *et al.* (2009) Review of the results from the International *C. elegans* First Experiment (ICE-FIRST). *Adv. Space Res.* 44, 210–216
70 Shmookler Reis, R.J. *et al.* (2006) Quantitative trait loci define genes and pathways underlying genetic variation in longevity. *Exp. Gerontol.* 41, 1046–1054
71 Chandler, C.H. (2010) Cryptic intraspecific variation in sex determination in *Caenorhabditis elegans* revealed by mutations. *Heredity* 105, 473–482
72 Elvin, M. *et al.* (2011) A fitness assay for comparing RNAi effects across multiple *C. elegans* genotypes. *BMC Genomics* 12, 510
73 Gaertner, B.E. *et al.* (2012) More than the sum of its parts: a complex epistatic network underlies natural variation in thermal preference behavior in *Caenorhabditis elegans*. *Genetics* 192, 1533–1542
74 Francesconi, M. and Lehner, B. (2014) The effects of genetic variation on gene expression dynamics during development. *Nature* 505, 208–211
75 Snoek, L.B. *et al.* (2014) Widespread genomic incompatibilities in Caenorhabditis elegans. *G3 (Bethesda)* 4, 1813–1823