DATABASE
The Journal of Biological Databases and Curation

Original article

WormQTL2: an interactive platform for systems genetics in *Caenorhabditis elegans*

Basten L. Snoek[1,2,†]*, Mark G. Sterken[1,†]*, Margi Hartanto[1,3], Albert-Jan van Zuilichem[1], Jan E. Kammenga[1], Dick de Ridder[3] and Harm Nijveen[3,]*

[1]Laboratory of Nematology, Wageningen University, Droevendaalsesteeg 1, NL-6708 PB Wageningen, The Netherlands, [2]Theoretical Biology and Bioinformatics, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands and [3]Bioinformatics Group, Wageningen University, Droevendaalsesteeg 1, NL-6708 PB Wageningen, The Netherlands

*Corresponding author: Basten L. Snoek. E-mail: l.b.snoek@uu.nl tel: +31 302 532 632. Correspondence may also be addressed to Mark G. Sterken. tel: +31 317 482 998; fax: +31 317 418 094 E-mail: mark.sterken@wur.nl and Harm Nijveen. tel: +31 317 484 706 fax: +31 317 418 094 E-mail: harm.nijveen@wur.nl

[†]These authors contributed equally to this work.

Abstract

Quantitative genetics provides the tools for linking polymorphic loci to trait variation. Linkage analysis of gene expression is an established and widely applied method, leading to the identification of expression quantitative trait loci (eQTLs). (e)QTL detection facilitates the identification and understanding of the underlying molecular components and pathways, yet (e)QTL data access and mining often is a bottleneck. Here, we present WormQTL2, a database and platform for comparative investigations and meta-analyses of published (e)QTL data sets in the model nematode worm *C. elegans*. WormQTL2 integrates six eQTL studies spanning 11 conditions as well as over 1000 traits from 32 studies and allows experimental results to be compared, reused and extended upon to guide further experiments and conduct systems-genetic analyses. For example, one can easily screen a locus for specific cis-eQTLs that could be linked to variation in other traits, detect gene-by-environment interactions by comparing eQTLs under different conditions, or find correlations between QTL profiles of classical traits and gene expression. WormQTL2 makes data on natural variation in *C. elegans* and the identified QTLs interactively accessible, allowing studies beyond the original publications.

Database URL: www.bioinformatics.nl/WormQTL2/

Introduction

The nematode *Caenorhabditis elegans* has been instrumental as a model organism in studying genotype-phenotype relationships. Its genetic tractability in combination with a rapid life cycle and a large body of experimental data provides a powerful platform for investigating the genetics of

complex traits. Extensive molecular, cellular and physiological insights have been obtained using knockout mutants, RNAi-treatment and various other techniques carried out in the canonical genotype Bristol N2 (1). Yet, due to many selection bottlenecks in the laboratory, the Bristol N2 strain has become the 'lab worm' and is not representative of the effect of wild-type alleles, as reviewed by (1). Over the last decade, the study of natural variation in *C. elegans* has made rapid progress, leading to the identification of over a dozen allelic variants contributing to natural phenotypic variation (2–26). Most quantitative genetics studies in this model animal have been conducted in recombinant inbred line (RIL) panels derived from crosses between the N2 laboratory strain (originally isolated in Bristol, UK) and the genetically divergent wild-isolate CB4856 (isolated in Hawaii, USA) (1, 27–29). A substantial amount of phenotypic, genotypic and high-throughput molecular data has been gathered across these recombinant inbred panels, as well as from introgression lines (30, 31), and many other wild isolates (32–34); for a more detailed overview, see reviews (1, 35). Together, the field of quantitative genetics in *C. elegans* has been very productive. However, accessing all this data and using it for follow-up studies or for comparative analysis can be challenging. Furthermore, recent in-depth studies on the ecology of *C. elegans* yielded even more phenotypic and genotypic information on novel wild isolates (34, 36–46). Inclusion of diverse genetic backgrounds in experiments is therefore a welcome and useful addition to the work on the canonical N2 strain as it can identify novel modifiers of well-studied pathways and thereby shed light on molecular mechanisms of genetic variation (46–52). A collection of data on wild isolates, including genome sequences, is curated and available via the *C. elegans* Natural Diversity Resource (CeNDR) (33).

The genetics of complex traits can be unraveled by performing quantitative trait locus (QTL) analysis. QTLs are parts of the genome that harbor genetic variation associated with trait variation measured between different genotypes. Most QTL studies in *C. elegans* make use of RILs derived from Bristol N2 and the genetically divergent Hawaiian strain CB4856 (28, 29). Traits such as body size, fecundity, aging or pathogen sensitivity have been linked to underlying loci by QTL analysis (11, 53–59). Moreover, gene expression studies comparing N2 and CB4856 show ample genotype-dependent gene expression variation (60, 61). When microarray platforms became affordable and more easily usable, as currently has happened for RNA-seq, the range of phenotypes in RIL populations was extended with genome-wide gene expression. The identified expression QTLs (eQTLs) are—like classical QTLs—polymorphic loci linked to gene expression variation (62, 63). eQTLs can be *cis* or *trans*-acting: a *cis*-eQTL is normally defined as

an eQTL mapping near the genomic location of the gene which it affects (usually within 1–2 Mb for *C. elegans*) (28, 31, 50, 64–67), while for a *trans*-eQTL genetic variation in another genomic region causes the change in gene expression. Both *cis*- and *trans*-eQTLs can shed light on the regulatory mechanisms underlying variation in molecular and phenotypic traits. Furthermore, the co-localization of *trans*-eQTLs, coined *trans*-bands or eQTL hotspots makes them interesting for regulatory network analysis, as it is assumed that one or a few polymorphic 'master regulator' genes affect the expression of many target genes. In principle, by connecting each gene to its regulators, gene regulatory networks can be constructed from these eQTLs (29, 68–75).

The identification of causal genes underlying *trans*-bands in *C. elegans* can be challenging. One of the complicating aspects is that the ultimate causal variant may act indirectly (e.g. through behavior or hormone), rather than via a direct route (e.g. a transcription factor) (3). To date, only two such causal variants have been experimentally confirmed: *npr-1* and *amx-2* (3, 50). Another limitation is the still relatively low resolution of current eQTL analysis, typically yielding eQTLs spanning large genomic regions with hundreds of genes. Therefore, combining data from different experiments will result in better contextual information leading to a more detailed reconstruction of the regulatory mechanisms and more specific candidate gene lists.

Here, we present WormQTL2 (www.bioinformatics. nl/WormQTL2/), the platform for systems genetics in *C. elegans*. WormQTL2 is based on a versatile and interactive analysis platform for *Arabidopsis* QTL data called AraQTL, www.bioinformatics.nl/AraQTL (76). We used this framework combined with the ideas from WormQTL (www.WormQTL.org), (77–79) to present the *C. elegans* QTL data in an interactive manner. The WormQTL2 interface improves on WormQTL (77–79), allowing for dynamic and interactive cross-study analyses to aid hypothesis driven genotype-phenotype investigations. Through WormQTL2, data on natural variation in *C. elegans* and the identified QTLs have been made accessible and interactively approachable, beyond the original publications. The majority of the data in WormQTL2 are expression QTLs based on data from six different *C. elegans* eQTL studies (Table 1) (28, 50, 64–67). Moreover, phenotypic QTLs are well represented with data on more than 1000 traits from 32 studies. This extends the exploratory options by integrating QTL data on classical phenotypes, gene expression levels as well as protein- and metabolite levels (2, 3, 13, 23, 30, 34, 41, 47, 49, 53–59, 61, 64, 80–90) (Table 2). In this paper, we present WormQTL2 and showcase its use by presenting short research scenarios.

Table 1. eQTL studies available in WormQTL2

Study	Population	Microarray type	eQTLs*	Stage	Environment
Li *et al.* (28)	N2 × CB4856 160 (2 × 80) RILs	Array 900HS Washington University	G, G × E	L3	16 or 24°C
Viñuela *et al.* (76)	N2 × CB4856 108 (3 × 36) RILs	Array 900HS Washington University	G, (G × E)	L4 (40 h), adult (96 h), old (214 h)	24°C
Rockman *et al.* (77)	N2 × CB4856 208 RIAILs	Agilent-015061 4 × 44K (G2519F)	G	Young adult (60 h)	20°C
Li *et al.* (78)	N2 × CB4856 60 RILs	Affymetrix 1.0 C. *elegans* tiling	G	Late L3	24°C
Snoek *et al.* (79)	N2 × CB4856 144 (3 × 48) RILs	Agilent 4 × 44K (V2)	G, (G × E)	L4	Control (48 h at 20°C) Heat-shock (46–2 h at 35°C) Recovery (46–2 h at 35°C–2 h at 20°C)
Sterken *et al.* (50)	MT2124 × CB4856 33 miRILs	Agilent 4 × 44K (V2)	G	Adult (72 h)	20°C

*eQTL G, genetic; eQTL GxE, environment specific eQTLs.

Results

eQTL studies in WormQTL2

WormQTL2 is a browser-based interactive platform and database for investigating expression and other QTL studies conducted in *C. elegans* (Figure 1). It enables access to the mapping data of six previously published eQTL studies (Table 1) (28, 50, 64–67). Together, these studies cover over 700 samples, including expression measurements of ∼20 000 different genes across different life stages and environmental conditions. The effect of genetic variation on gene expression is presented in 11 genome-wide sets of eQTLs from three different RIL populations. The first is a CB4856 × N2 RIL population (28). The second a CB4856 × N2 recombinant inbred advanced intercross line (RIAIL) population (91, 92). The third a mutation introgressed RIL population resulting from a cross between a *let-60* gain-of-function mutant in an N2 background, MT2124, with CB4856 (11, 50). For the Li *et al.* (28), Viñuela *et al.* (64) and Li *et al.* (66) studies, the eQTLs were re-mapped with the most recent genetic maps used in Snoek *et al.* (65), which can be obtained from WormQTL2 at the download page accessible by pressing the 'Download' button (27, 65).

The first eQTL study in *C. elegans* was published in 2006 by Li *et al.* (28), where variation in gene expression was reported among N2 × CB4856 RILs grown at two different temperatures (16 vs 24°C). In 2010, three eQTL studies in *C. elegans* were published (64, 66, 67). Viñuela *et al.* showed age specific eQTLs, Li *et al.* investigated variation in splice variants and Rockman *et al.* used eQTLs to show

that phenotypic variation in *C. elegans* is determined by selection on linked sites. These three eQTL studies were used in many follow-up investigations/analyses focusing on how genetic variation affects gene expression, on the regulation of specific genes or on the molecular pathways underlying phenotypic variation. For instance, a *trans*-band on chromosome X observed by Rockman *et al.* was later identified to result from mild starvation and linked to genetic variation in the *npr-1* gene (3). Snoek *et al.* showed the effect of heat-stress and recovery on eQTL distribution and occurrence as well as the contribution of *trans*-eQTLs to cryptic variation (65), and Sterken *et al.* showed the interaction between genetic variation, gene expression and a *let-60* gain-of-function mutation (11, 50). An important overall conclusion drawn from these analyses was that eQTLs are highly dependent on the ambient environment and sensitive to induced background mutations.

Lately, the diversity of molecular phenotypes for which natural variation can be found and used to map QTLs has been expanded to proteins (80, 93) and metabolites (81). The associations of these molecular phenotypes with variation in gene expression, eQTLs and classical phenotypes and QTLs have yet to be explored. For such applications, WormQTL2 provides the data and the interactive platform.

QTL studies in WormQTL2

WormQTL2 currently provides access to the data and QTLs of 32 RIL-based QTL studies in *C. elegans* (Table 2). These studies include many 'classical phenotypes' as well

Table 2. QTL studies available in WormQTL2

Paper reference	Title	Parental lines	Type	Traits	QTLs $-\log_{10}(p) > 3.5$
Andersen *et al.* (3)	A variant in the neuropeptide receptor *npr-1* is a major determinant of *C. elegans* growth and physiology.	N2 × CB4856	RIAIL	7	10
Andersen *et al.* (82)	A powerful new quantitative genetics platform, combining *C. elegans* high-throughput fitness assays with a large collection of recombinant strains.	N2 × CB4856	RIAIL[1]	34	27
Balla *et al.* (41)	A wild *C. elegans* strain has enhanced epithelial immunity to a natural microsporidian parasite	N2 × CB4856	RIAIL	1	0
Bendesky *et al.* (8)	Catecholamine receptor polymorphisms affect decision-making in *C. elegans*.	N2 × CB4856	RIAIL	1	2
Bendesky *et al.* (7)	Long-range regulatory polymorphisms affecting a GABA receptor constitute a QTL for social behavior in *C. elegans*.	N2 × CB4856	RIAIL	2	1
Evans *et al.* (83)	Shared genomic regions underlie natural variation in diverse toxin responses.	N2 × CB4856	RIAIL[1]	384	590
Duveau and Felix (47)	Role of pleiotropy in the evolution of a cryptic developmental variation in *C. elegans*.	N2 × AB1	RIL	3	3
Elvin *et al.* (49)	A fitness assay for comparing RNAi effects across multiple *C. elegans* genotypes.	N2 × CB4856	RIL	108	20
Gaertner *et al.* (84)	More than the sum of its parts: a complex epistatic network underlies natural variation in thermal preference behavior in *C. elegans*.	N2 × CB4856	RIAIL	5	5
Gao *et al.* (81)	Natural genetic variation in *C. elegans* identified genomic loci controlling metabolite levels.	N2 × CB4856	RIL	378	154
Glater *et al.* (85)	Multigenic natural variation underlies *C. elegans* olfactory preference for the bacterial pathogen *Serratia marcescens*.	N2 × CB4856	RIAIL	2	2
Greene *et al.* (86)	Regulatory changes in two chemoreceptor genes contribute to a *C. elegans* QTL for foraging behavior.	MY14 × CX12311	RIL[2]	2	1
Gutteling *et al.* (58)	Environmental influence on the genetic correlations between life-history traits in *C. elegans*.	N2 × CB4856	RIL	6	2
Gutteling *et al.* (59)	Mapping phenotypic plasticity and genotype-environment interactions affecting life-history traits in *C. elegans*.	N2 × CB4856	RIL	18	5
Harvey *et al.* (88)	Quantitative genetic analysis of life-history traits of *C. elegans* in stressful environments.	N2 × DR1350	RIL	26	4
Harvey (87)	Non-dauer larval dispersal in *C. elegans*.	N2 × DR1350	RIL	2	2
Kammenga *et al.* (5)	A *C. elegans* wild-type defies the temperature-size rule owing to a single nucleotide polymorphism in *tra-3*.	N2 × CB4856	RIL	1	1
Large *et al.* (6)	Selection on a subunit of the nurf chromatin remodeler modifies life history traits in a domesticated strain of *C. elegans*.	LSJ2 × CX12311	RIL	5	7
Lee *et al.* (89)	The genetic basis of natural variation in a phoretic behavior.	N2 × CB4856	RIAIL1	1	1
McGrath *et al.* (13)	Quantitative mapping of a digenic behavioral trait implicates globin variation in *C. elegans* sensory behaviors.	N2 × CB4856	RIAIL	2	4

.   Continued

Table 2. QTL studies available in WormQTL2

Paper reference	Title	Parental lines	Type	Traits	QTLs $-\log_{10}(p) > 3.5$
Nakad *et al.* (53)	Contrasting invertebrate immune defense behaviors caused by a single gene, the *C. elegans* neuropeptide receptor gene *npr-1*.	N2 × CB4856	RIL	24	17
Noble *et al.* (2)	Natural variation in *plep-1* causes male–male copulatory behavior in *C. elegans*.	QG5 × QX1199	RIL[3]	5	5
Rockman and Kruglyak (91)	Recombinational landscape and population genomics of *C. elegans*.	N2 × CB4856	RIAIL	7	6
Rodriguez *et al.* (57)	Genetic variation for stress-response hormesis in *C. elegans* lifespan.	N2 × CB4856	RIL	9	1
Schmid *et al.* (11)	Systemic regulation of RAS/MAPK signaling by the serotonin metabolite 5-HIAA	MT2124 × CB4856	RIL[4]	2	3
Singh *et al.* (80)	Natural genetic variation influences protein abundances in *C. elegans* developmental signaling pathways.	N2 × CB4856	RIL	10	1
Snoek *et al.* (34)	A multi-parent RIL population of *C. elegans* allows identification of novel QTLs for complex life history traits.	JU1511 × JU1926 × JU1931 × JU1941	mpRIL[5]	21	72
Snoek *et al.* (55)	Widespread genomic incompatibilities in *C. elegans*.	N2 × CB4856	RIL	4	3
Stastna *et al.* (54)	Genotype-dependent lifespan effects in peptone deprived *C. elegans*.	N2 × CB4856	RIL	2	2
Viñuela *et al.* (76)	Genome-wide gene expression regulation as a function of genotype and age in *C. elegans*.	N2 × CB4856	RIL	2	0
Zdraljevic *et al.* (23)	Natural variation in a single amino acid substitution underlies physiological responses to topoisomerase II poisons.	N2 × CB4856	RIAIL[1]	2	3
Zhu *et al.* (90)	Compatibility between mitochondrial and nuclear genomes correlates with the quantitative trait of lifespan in *C. elegans*.	N2 × CB4856	RIAIL	15	11

[1] Supplemented by an QX1430 × CB4856 cross.
[2] CX12311 is N2 without *npr-1* and *glb-5*.
[3] QG5 is *him-5* (e1490) > AB2, QX1199 is *him-5* (e1490) > CB4856.
[4] MT2124 is a *let-60* gain-of-function mutant.
[5] Multi (four) parental cross.

as molecular phenotypes such as metabolite and protein levels. We compiled the list of studies using two reviews listing older studies, mostly pre-2000 (1, 35), and performed a literature search for more recent studies from 2000-onwards. For many studies, we could obtain the relevant data from the supplemental information (2, 3, 5, 6, 21, 34, 41, 47, 49, 53, 54, 57–59, 64, 80, 81, 83–86, 89–91). Where such data were not provided in journal supplements, approaching the authors was successful for most studies (6–8, 11, 13, 55, 82, 87, 88, 91). Data from Bergerac BO × N2 populations were not included as the genetic map only consisted of a few markers and the Bergerac BO strain contained active transposons, complicating the interpretation of the data (35). In summary, WormQTL2 provides full access to all QTL studies up to 2018, where the required data were available or kindly provided upon request.

The data sets included to provide full access to all underlying raw data. For each study, a genetic map is available (updated to genome version WS258 coordinates), as well as the raw data used for mapping and the output of a single marker model. These data allow users to either access the raw data and run alternative analyses as they wish or access already mapped QTL information. All data were mapped using a single marker model, which is shown in WormQTL2. In total, 929 QTLs were mapped for 1091 traits ($-\log_{10}(p) > 3.5$). To make these data insightful, trait-names were standardized where possible (e.g. consistent use of 'body' in relation to measurements of size, volume and length of the animal body). Furthermore, traits were coupled to gene ontology (GO) terms, to facilitate coupling to transcriptomics data. These curatorial steps greatly facilitate analytical access, increasing the ways in which the user can interact with the data.

Figure 1. WormQTL2 Homepage. On the top of the page, the navigation bar can be found. This includes the WormQTL2 logo, which functions as a home button. It also includes a fast link to the Correlation and Locus overviews as well as links for help, data download and visual examples. The search box is located in the center, in which genes, phenotypes and GO terms can be entered. Shown in the blue middle square are the buttons for the investigations of single traits, correlating QTL profiles, QTLs at a specific locus, all eQTLs of an experiment, eQTLs per GO term and a list of all phenotypes for which QTLs can be shown. On the right side, a bar with information about WormQTL2 and QTLs in general can be found.

Starting your search using WormQTL2

The homepage offers several approaches to investigate QTL data, including searching for individual traits, correlating QTL patterns and finding traits that have a QTL at a specific locus (Figure 1). Furthermore, six options are provided for quick navigation to specific investigation paths. A detailed description of WormQTL2 navigation and use can be found in the manual (Supplementary Manual). In general, the search function can be used to find QTL profiles for one or more traits or genes in one or all experiments. Also, GO terms can be entered to find all genes annotated with that GO term and investigate their eQTL profiles. Any search input not directly matching with a gene ID or GO term will report the genes and GO terms with matching descriptions. Divided over several interactively linked pages, different functions are available for investigation and exploration.

Selecting experiments

Experiments of interest can be selected from the 'Experiment overview' page. In this table, basic information about the experiment can be found, such as population used, developmental stage, temperature and source publication. Publications are linked to their PubMed pages for easy access to the experimental details. Data from all experiments, such as QTL profiles, genetic maps and phenotypes,

are available in WormQTL2 and can be downloaded or directly accessed in flat text format, for instance to further explore with programming languages such as R or Python.

For easy access of the main functionality, every WormQTL2 page shows the navigation bar at the top of the page (Figure 1). It can be used for a selection of graphical overviews, investigations and information. To return to the homepage and search function, the 'WormQTL2' button in the left upper corner can be used; for each data set, 'correlation' can be used to find traits with correlated QTL profiles; 'locus' shows all traits with a QTL at the specified marker or genomic position; and frequently asked questions and other info can be found by clicking 'help'. 'Examples' leads to an interactive graphical overview of several different functions of WormQTL2.

WormQTL2 use cases WormQTL2 was developed to facilitate meta analyses of QTL and eQTL data for extended investigations and distinguishes itself from other databases by enabling interactive selection of groups of genes or traits based on a common genetic effect. Physical marker positions have been used to integrate the genetic maps of the different populations to enable direct comparisons between eQTLs found in different experiments and populations. Furthermore, WormQTL2 uniquely allows users to find whether a group of genes (such as those sharing a specific GO term) have a shared *trans*-band, a so-called hotspot of
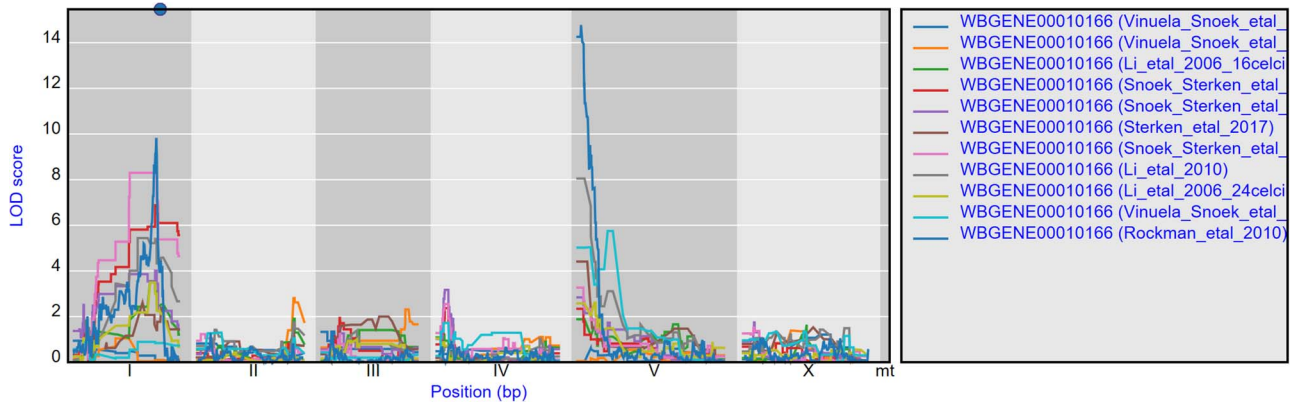
Figure 2. eQTL profiles of gmd-2 from all experiments hosted in WormQTL2. The significance profile ($-\log10(p)$) per experiment for gmd-2 is shown in different colors per experiment. The legend is shown on the right and shows the colors for the different experiments. Chromosomes are indicated by different gray backgrounds and below the *x*-axis. The web-based plot is interactive; a mouse over provides the exact base pair position of each QTL or other point on the profile.

eQTLs, which can then be efficiently investigated further for identifying other genes with co-locating eQTLs by the integrative tools. The genes with co-locating eQTLs can then be exported as a list for further investigations or to an external analysis platform. Overall, genes with a shared genetic architecture are easily investigated within and outside WormQTL2, making it a versatile tool for *C. elegans* researchers.

Example 1: cryptic variation in gene expression

The environment-specific as well as environment-independent eQTLs for individual or small groups of genes can be easily found by using the search box on the homepage. For instance, gene *gmd-2* has very similar eQTL profiles across experiments, a *cis*-eQTL at chromosome I and a *trans*-eQTL at chromosome V, most prominent in the juvenile and young adult stages (Figure 2) (28, 64, 65, 67). But both *cis*- and *trans*-eQTL are absent from older worms (64, 67) as well as when the genetic background contains a *let-60gf* mutation (50). This shows the hidden/cryptic variation affecting the expression levels of a gene across experiments. This can be very specific, for example *hsp-12.3* has a *cis*-eQTL on chromosome IV in Rockman *et al*. (67) and not in any of the other experiments, yet it has a co-locating *trans*-eQTL in both heat stress and recovery conditions in Snoek *et al*. (65).

Similar cryptic variation can also be investigated for other genes, such as *pgp-7*, which has a very prominent *cis*-eQTL on chromosome X when a *let-60gf* mutation is present in the genetic background (50). Yet in many other experiments, it has a *trans*-eQTL on chromosome V (64, 65, 67). This shows that different polymorphic regulators exist whose action depends on the developmental stage and on the genetic background. The most significant *trans*-eQTL

was found in the control conditions of Snoek *et al*. (65) (Table 1); by using WormQTL2's correlation function on this experiment, we can find two *pgp-7* homologs, *pgp-5* and *pgp-6*, which also have a *trans*-eQTL at this locus and likely share a common polymorphic regulator (Figure 3).

The correlation between eQTL profiles can show co-regulated groups of genes, the experiment in which they are co-regulated and the regulatory loci involved. When we inspect *daf-18* and the genes with a correlated eQTL profile, we find that *daf-18* is part of a group of co-regulated genes with two regulatory loci only during heat-stress conditions (65), one on chromosome I and one on chromosome V (Supplementary Figure S1). Moreover, when the group of genes is enriched for one or more GO terms, a table is provided below the gene table. The *daf-18* co-regulated genes are enriched for larval development, suggesting an effect of heat stress on larval development through *daf-18* expression variation and possibly the loci on chromosome I and chromosome V. Comparing the eQTL profiles of genes and groups of genes in different experiments shows the dynamic nature of polymorphic regulatory loci and the genetic architecture underlying cryptic variation.

Example 2: GO term investigation

Groups of genes can also be selected based on GO terms. Per experiment, the eQTL profiles of all genes annotated with a specific GO term can be shown, with the most significant 15 pre-selected. When, for example, 'cell cycle' is entered in the search box, a list of genes and GO terms is returned. From this list, we can pick GO term 'regulation of cell cycle' and study the eQTL profiles of the genes involved in this process (Figure 4). In the Viñuela *et al*. (64) juvenile set, 39 genes with co-locating eQTLs can be found on chromosome I, indicating a polymorphic regulator for the cell cycle can be
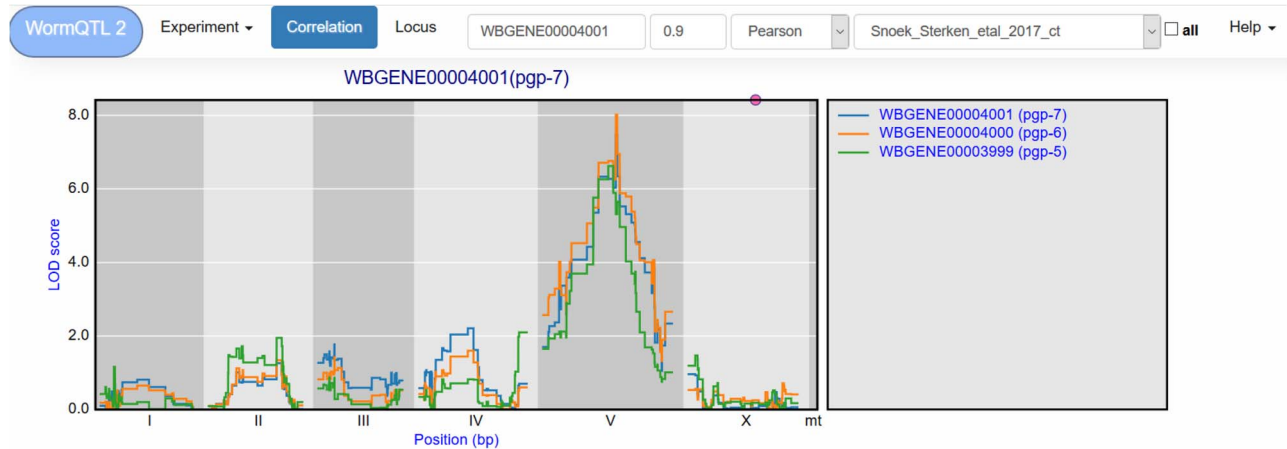
Figure 3. Correlation between eQTL profiles of *pgp-7*, *pgp-6* and *pgp-5*. Correlation between the eQTL profile of *pgp-7* and the other genes was set to report those profiles with a Pearson correlation >0.9, see navigation bar at the top. Genes are shown in different colors. The legend is shown on the right. Chromosomes are indicated by different gray backgrounds and below the *x*-axis. The web-based plot is interactive; a mouse over provides the exact base pair position of each QTL or other point on the profile.

found at this locus. When we observed the eQTL profiles in other studies, the co-location at the locus is gone, indicating the regulator is specific to the juvenile life stage.

This can also be observed for the genes annotated with a related GO term 'chromosome segregation' where 41 eQTLs can be found co-locating on chromosome I in the juvenile stage, but not in the reproductive or old stage. In the old stage, 14 co-locating eQTLs can be found at chromosome V (Supplementary Figure S1).

These examples show that starting with groups of genes sharing a GO term can be a great start for exploring eQTL data in order to find co-locating eQTLs of genes with a shared function and possibly identify the position of GO term specific polymorphic regulators.

Example 3: exploring a trans-band

A *trans*-band can be selected by clicking the histogram under the *cis/trans* plot. This leads to a list of genes that pass the significance threshold at the selected marker. For example, in the Rockman *et al.* (67) data, marker rmm1258 (chromosome X at 3.8 Mb) can be selected, leading to a list of 126 genes at a $-\log_{10}(p)$ threshold of 3.5. However, this list contains both *cis*- and *trans*-eQTLs. To select the *trans*-eQTLs, a minimal distance threshold can be specified to remove genes that are located close to the selected locus. For example, setting this threshold to 2 million base pairs narrows-down the list to 111 genes (Figure 5). These can be investigated further, both within and outside of WormQTL2, to predict their regulator pathway or biological function.

WormQTL2 offers users the option to investigate the eQTL patterns of different studies at the locus location. For instance, it can be informative to determine if a *trans*-band

occurs in other studies, by selecting a study from the drop-down menu. WormQTL2 will select and display the eQTLs at the nearest marker for any study selected. When applied to the rmm1258 *trans*-band, by selecting the heat-shock condition of the Snoek *et al.* (65) study, we find no clear *trans*-band at the corresponding location (PredX3820001), only six genes with a *trans*-eQTL. However, in the control condition of this study 26 genes are found to have an eQTL at this position. This is a clear demonstration that this *trans*-band can be environment specific and disappears under heat-shock in L4 stage nematodes.

Beyond WormQTL2

Using the data stored, explored and selected through WormQTL2, there are several options for further analysis. Using the 'Download selected trait IDs' button, a list of WormBase IDs can easily be selected and copied. These IDs can subsequently be used in other online *C. elegans* resources, such as the Serial Pattern of Expression Levels Locator [SPELL (94)]. For example, the 111 wormbase-IDs from the *trans*-eQTLs can be inserted as query to find gene expression datasets in which the queried genes display co-expression. Visual inspection of the hits identifies three studies with treatment-related variation in this set of genes. One is the original study (67), one is a study on the innate immunity of *C. elegans* and *Pristionchus pacificus* (95) and, most interestingly, a study on an *aptf-1* mutant (96). This mutant shows lower expression of *flp-11*, suggesting that the *trans*-band is somehow related to neuronal activity. From literature, we know this is actually the case, as the underlying gene is *npr-1*, for which a variant (215V in N2) has a neomorphic gain-of-function mutation, making it responsive to both FLP-21 and FLP-18 [reviewed by
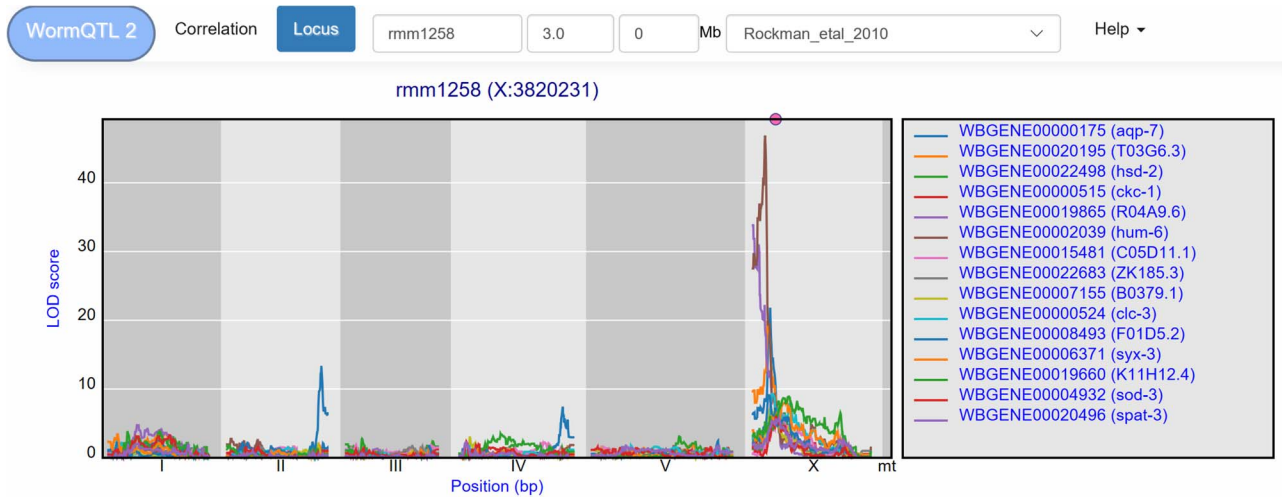
Figure 4. Co-locating eQTL profiles of genes annotated with GO term 'regulation of cell cycle'. The eQTL profiles of the top-15 highest eQTLs are shown as profiles. A histogram of all peaks $-\log10(p) > 3.5$ is shown below the profile plot to detect possible regulatory hotspots, as can be observed here on chromosome I. Genes are shown in different colors. The legend is shown on the right. Chromosomes are indicated by different gray backgrounds and below the *x*-axis. The web-based plot is interactive; a mouse over provides the exact base pair position of each QTL or other point on the profile.
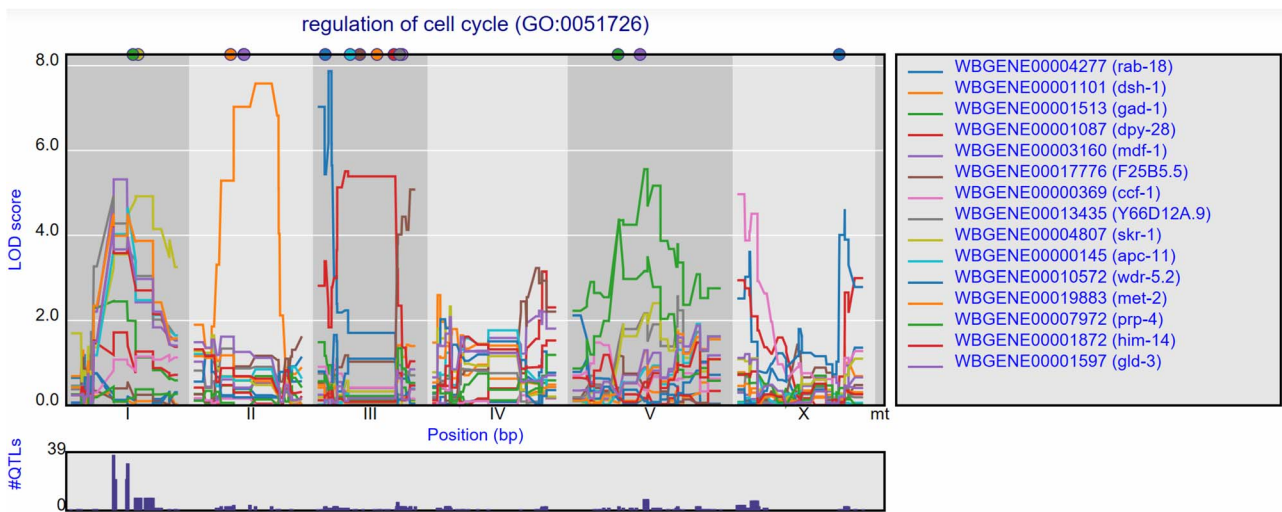

Figure 5. eQTL transband on chromosome X in Rockman *et al.* (67). The eQTL profiles of the top-15 highest eQTLs at marker rmm1258 in Rockman *et al.* are shown as profiles. Colors indicate different genes. The legend is shown on the right. Chromosomes are indicated by different gray backgrounds and below the *x*-axis. The web-based plot is interactive; a mouse over provides the exact base pair position of each QTL or other point on the profile. Moreover, more than the top-15 can be added to the plot creating a custom overview.

(1)]. There are many other databases that can be consulted similarly for further investigation, such as WormNet (97), WormBase (98), MODENCODE (99)/MODERN (100), Genemania (101) and StringDB (102).

Discussion

The WormQTL2 platform for data access

WormQTL2 offers a comprehensive and interactive QTL-data platform for *C. elegans*. It complements and extends existing data analysis and presentation platforms for QTL studies in *C. elegans* such as WormQTL [www.WormQTL.org (78, 79)] and WormQTL-HD [www.WormQTL-HD.org (77)] and is developed to support *C. elegans* investigators in the analyses of natural genetic variation. All data in WormQTL2 are cross-linked, which allows for the comparative investigation of QTL patterns across studies and phenotypes in a user-friendly interactive way. This can be of great aid as often, published (e)QTL studies do not include the complete QTL profiles. Several years ago, we developed the WormQTL platform to serve as a central repository for these QTL profiles, including the data needed for re-mapping. However, WormQTL lacks interactive tools

suitable for further analyses of eQTL profiles, limiting its practical use for direct exploration. In WormQTL2, QTL profiles for genes, metabolites and phenotypes can be viewed and studied interactively. This facilitates the integration of different data sets and allows for comparisons which would otherwise be cumbersome and laborious.

WormQTL2 offers access to RIL-based (e)QTL-data in *C. elegans*. Currently, WormQTL2 offers access to all published eQTL data sets and a majority of the published phenotypic QTL data sets. All data sets were curated and the eQTL studies and genetic maps were updated to a recent *C. elegans* genome version (currently the database runs based on WS258). The QTLs of phenotypes of the included studies were re-mapped using a single marker model for uniformity. This does lead to some differences compared with the original study if specific models and mapping procedures were used. However, next to the interactive front-end, the platform also offers access to the raw data, allowing more experienced users to download the data and run custom investigations, extending the reusability of the observations.

The WormQTL2 platform currently limits itself to RIL-based (e)QTL data. Future development will first focus on integrating data from introgression-line (IL)-based studies. There is a rich body of published studies utilizing introgression lines for QTL validation, but also as alternative to RIL-based genome-wide studies. Especially, the genome-wide N2 × CB4856 IL panel and a set of chromosome-substitution lines have been used (30, 103). Currently, the *C. elegans* field increasingly makes use of Genome-Wide Association Studies (GWAS) and wild isolates (32, 33, 40, 91). WormQTL2 is currently developing links to genetic variants in QTL regions through the CeNDR (33).

Data exploration and analysis through WormQTL2

WormQTL2 offers the user the possibility to compare (e)QTL patterns across studies. From literature across model species, it is currently clear that *cis*-eQTLs: (i) can result from hybridization differences when using microarrays (67, 104–106); (ii) explain more variance than single *trans*-eQTL (67, 107); (iii) are constitutively found across experiments using the same populations and environments (28, 64–66, 75, 108); and (iv) are often found for polymorphic genes (28, 67, 69, 75). In contrast, *trans*-eQTLs are strongly environment dependent and seem in large part unique across environments (28, 64, 75, 109).

Comparative analysis of eQTL profiles does come with some inherent limitations. The main limitation is platform-based. WormQTL2 offers access to eQTL studies from four different microarray platforms. As microarray technology uses probes on a glass-slide, false negatives may occur due to the absence of a probe, preventing the interrogation for that gene (67, 104). Closely related genes can cross-hybridize due to probe similarity (although we try to minimize the risk by excluding probes with multiple blast-hits). Furthermore, false positive eQTLs could be obtained when hybridisation differences due to sequence polymorphisms are mistaken for transcript abundance variation. These QTLs, however, can be used as genetic markers or to detect wrongly labeled samples (105, 106, 110). Hence, users should be mindful about technical limitations when comparing results from different experiments. Nevertheless, for most eQTLs and general patterns, cross-platform comparisons can be insightful and useful (50).

Comparison of (e)QTL studies through WormQTL2 depends on the mapping populations involved. We offer analytical access to studies widely differing in statistical power and RIL population. For example, the number of RILs used per (e)QTL data set ranges from 36 for each of the three conditions in Viñuela *et al.* (64), up to 200 in Rockman *et al.* (67) and Snoek *et al.* (34) or over 500 in the CeMEE panel of Noble *et al.* (111). Furthermore, the size of the genetic map (in centimorgans) of the N2 × CB4856 RIAIL populations is larger than the N2 × CB4856 RIL population (27, 28, 67). Genetic maps of the mutation included RIL populations and the N2 × DR1350 populations include areas without genetic variation (or information on the genotype) (11, 50, 88), whereas a multi-parental RIL panel contains multiple SNP distribution patterns (34). Finally, the number of markers used to genotype RILs is different between mapping populations. In WormQTL2, we therefore present the eQTL profiles of re-mapped studies so that the data sets are directly comparable.

Future developments WormQTL2 aims to provide re-mapped (e)QTLs in *C. elegans*. Currently, re-mapping has been done using a single marker model, making the output comparable across studies. As all the relevant data for mapping are hosted, it is possible to integrate alternative models or integrate analysis of different experiments in one (e)QTL mapping model. Genetic maps can also be improved by including gene expression markers (50, 65, 105, 106, 110), through sequencing (27) or use of RNA-seq (34, 112). This will lead to eQTLs with a higher resolution and better regulatory prediction. Easy access to the data already enables an efficient start for further exploiting eQTLs and other system genetics data by anyone. In future updates of WormQTL2, these QTL mapping functions can be implemented.

Combining established high-throughput measurement techniques such as next-generation sequencing (27, 32–34), proteomics (80, 93), metabolomics (81) and phenomics (82, 83, 113) offers great potential for further quantitative

genetic analyses across different levels. This wealth of data makes the storage, access and especially the generation of useful and meaningful connections within and between the different types of data increasingly important. Moreover, results from different types of mapping populations can be included. In this way, the advantages of IL populations (30, 54–56, 103, 114, 115), RIL populations (28, 50, 92), multi-parental mapping panels (34, 111) and sets of wild-isolates for GWAS (33) can be combined. With this in mind, the next steps for WormQTL2 will be linking eQTL data to polymorphisms from massive sequencing projects of many different ecotypes (32, 33) and including eQTLs and SNPs obtained from RNA-seq experiments. When stored in, and visualized from, the same platform, the SNPs and pheno-types enable the integration of QTL mapping and GWAS investigation, further increasing the detection power of both methods. For example, eQTL data sets have been success-fully combined with results from transcriptomic GWAS (116) and allele-specific expression RNA-seq experiments (109).

New tools for investigation and visualization will be developed in a modular fashion for easy integration and deployment within WormQTL2. Annotations can be expanded beyond GO terms, for example with pathway knowledge, e.g. as available through the Kyoto Encyclo-pedia of Genes and Genomes (KEGG; www.genome.jp/kegg/), or with gene association networks such as WormNet (97), StringDB (102) and Genemania (101). To further investigate the relation between genotype and phenotype variation, WormQTL2 will be expanded with published and new classical/phenotypic QTLs. This enables searching for the possible molecular components underlying a QTL for a specific phenotype and finding the causal genes. Combining highly detailed molecular data, such as generated and shared by the modENCODE consortium (99, 117, 118), such as transcription factor- and histone-binding sites or protein–protein interactions will allow for even more powerful analyses.

WormQTL2 has been designed to easily store and share upcoming RNA-seq data and eQTLs from this data, QTLs from metabolomics and proteomics, and visualize and ana-lyze these together. Comparing sets of genes through func-tional enrichment will enable an even better, more tar-geted, approach in candidate gene selection and network generation to link gene expression, genetic variation and function. In the near future, tools will be developed to investigate genetic variation in a more systematic, genome-and population-wide manner, enabling more complete and higher resolution system genetics.

We believe the (e)QTL data in WormQTL2 will greatly benefit the *C. elegans* research community, providing a rich source of genetic interactions specifically to worm

biologists and geneticists in general. WormQTL2 will serve as a solid platform for in-depth analysis of these interactions to help chart the *C. elegans* gene regulatory networks.

Experimental procedures

Transcriptome data

Data sets of the six eQTL experiments were retrieved from GEO or ArrayExpress: GSE5395 (28), GSE15778 (66), GSE17071 (64), GSE23857 (67), E-MTAB-5779 (65) and E-MTAB-5856 (50). The platform data were also obtained from GEO or ArrayExpress: GPL4043 (28, 64), GPL5634 (66), GPL7727 (67) and A-MEXP-2316 (50, 65). The microarray probes were re-mapped against the *C .elegans* reference genome version WS258 using blastn [version 2.6.0, win x64 (119)]. Probes with multiple high-ranking matches to different genes were censored.

Phenotypic data

The publications reporting *C. elegans* phenotypic QTLs were used to acquire phenotype and genotype data required for QTL mapping. This was done by taking the data directly from separated supplementary information or by contact-ing the authors. We curated data from 32 publications, comprising 1091 traits (Table 2). For each publication, the raw trait data per strain, the genetic map and the QTL data are made available.

Using the obtained phenotypes and the most updated genotype data, the QTLs for each study were re-mapped using a linear single marker model:

$$y_{i,j} \sim x_j + e_j,$$

where y is the phenotype value of RIL j based on the func-tion of marker genotype i. Subsequently, $-\log_{10}(P\text{-value})$ of 3.5 from each marker analysis was used as the threshold to determine the significant QTL shown in Table 2. This analysis was performed using a custom-made script in 'R' (version 3.4.4, win x64).

Genetic maps

For each population, the most detailed genetic map avail-able was used for remapping the eQTLs. For the experi-ments on the N2 × CB4856 RILs, this was a low-coverage sequencing-based genetic map (27) consisting of 729 mark-ers. As not all RILs in this population were sequenced, the genotypes of those RILs were imputed (65). For the N2 × CB4856 RIAIL population, a SNP-based genetic map consisting of 1454 markers was used (91). The genetic map of the MT2124 × CB4856 population consists of an

expanded FLP-map with 247 markers (11, 27, 50). The marker locations of each map were updated to the positions in reference genome version WS258.

For the phenotype QTL studies, the genetic map used in re-mapping can also be found on WormQTL2. This includes five additional genetic maps. First, the map for the expanded RIAIL set, which added 359 strains to the panel, without the N2 *npr-1* allele and a transposon insertion to reduce the effect of *peel-1* (82). Second, the JU605 × JU606 RILs, which are made from N2 with a *let-23(sy1)* mutation crossed with an AB1-genetic background introgression line with an N2-segment containing the *let-23* mutation (47). Third, the MY14 × CX12311 RILs, where CX12311 is an N2 strain with the wild-type *npr-1* and *glb-5* alleles (86). Fourth, RILs from an N2 × DR1350 cross (88). Fifth, RILs from a QG5 × QX1199 cross, where QG5 is the strain AB2 and QX1199 is CB4856, both carrying a *him-5(e1490)* mutation (2). Also, these maps were updated to WS258 coordinates.

Data analysis availability

The analytical scripts are available via https://git.wur.nl/published_papers/WormQTL2.

Microarray normalization and processing

For each array-type, recommended normalization methods were used. Each study was normalized independently (120). For all the normalization procedures, the limma-package from Bioconductor (121) was used in 'R' (version 3.4.2, win x64). The array data of the two studies based on the GPL4043 platform were background corrected using the *subtract* method, the within-array normalization method used was *printtiploess* and between-array normalization method used was *quantile*. The array data of the tiling-array (GPL5634) (66) was re-processed from the raw data and was batch corrected to remove between batch effects. Thereafter, the tiling array data were summarized per gene using the *quantile* function in 'R'. All five quantiles were used as input for subsequent analyses. For the two Agilent platforms (GPL7727 and A-MEXP-2316), no background correction was applied (122), the within-array normalization method used was *loess* and the between-array normalization method used was *quantile*.

After normalization, the intensities were $\log_2$-transformed for subsequent analyses. The Li *et al.* (28) experiment on the GPL4043 platform was corrected for array-specific effects due to a heterogeneous hybridization environment. In order to remove this effect, the difference between the array and the total average over all arrays was subtracted from the samples. The Viñuela *et al.* (64) experiment on the GPL4043 platform did not suffer from such an artifact.

Wrongly labeled samples

In order to reduce the noise in re-mapping, the correlation between the transcriptome profiles and the genotypes of the population used was determined based on known *cis*-eQTL in *C. elegans* using the method described in (65, 105). If switched labels were detected, these were corrected, and if no fitting correlations could be made, samples were censored. The data sets with correct strain labels and normalized values were made available for downloading and were used to produce the genotype split plots.

eQTL mapping

For each study, the eQTL were re-mapped using a single marker model, as in (50, 65). In short, the linear single marker model,

$$y_{i,j} \sim x_j + e_j,$$

was used, where the $\log_2$-normalized intensity (y) of spot i of RIL j was explained over the genotype at marker location x of RIL j. Per gene, the spot with the highest significance was selected as representative. The significance of the correlation and its effect were used to produce the *cis*/*trans* eQTL plots of the study and the detailed eQTL profile per gene.

Threshold determination

In WormQTL2, the user can set the threshold for determining if an eQTL is significant as this threshold can be dependent on many factors or even user preference. However, the default set thresholds are those used in the original papers: $-\log_{10}(p)$ thresholds for Li *et al.* (66); 4.2, Viñuela *et al.* (64); 3.8, Rockman *et al.* (67); 4.5, Snoek *et al.* (65); control 3.9, heat shock 3.9, recovery 3.9 and Sterken *et al.* (50); and 3.5 are used as the default settings.

QTL correlation analysis

All pairwise correlations between eQTL patterns were calculated using the Pearson correlation coefficients between the $-\log_{10}(p)$ values of the eQTL patterns of genes within an experiment using a custom python function (https://git.wageningenur.nl/nijve002/eleqtl/). Searching for genes with an eQTL at a specific locus is implemented by selecting genes that have a $-\log_{10}(p)$ score above the given threshold at the marker closest to the specified locus. To select *trans*-eQTLs, genes with a *cis*-eQTL can be excluded based on their physical distance to this marker.

Additional data and webpage development

GO terms were downloaded from WormBase and gene descriptions from Ensembl BioMart (123). WormQTL2

was developed using the Python Django web framework. The backend runs on an Ubuntu 17.10 Linux server, using the Apache web server version 2.4.27 and a MySQL 5.7.13 database. The web frontend is implemented via Django templates in HTML and Javascript, using the D3 library and Jquery. The *cis/trans* plot and QTL profile plots build upon work by Karl Broman (124).

Legacy data hosted at WormQTL

Transcriptomics and genomics data from three papers that were hosted on WormQTL, not hosted anywhere else and not falling under the category QTL experiments, were submitted to ArrayExpress. This concerns data from an experiment investigating genetic variation in 48 *C. elegans* isolates (E-MTAB-8126) and gene expression variation in these isolates (E-MTAB-8132) (38), transcriptomics data from an experiment comparing N2 and a *bar-1* loss-of-function mutant (EW15; ga80) (E-MTAB-8128) (125), and transcriptomics data from an experiment growing eight strains on two different bacteria (E-MTAB-8129) (126).

Supplementary Data

Supplementary data are available at *Database* online

Author contributions

H.N. and B.L.S. came up with the idea for WormQTL2. H.N., M.G.S. and B.L.S. designed WormQTL2. H.N. wrote the code for WormQTL2. M.G.S., B.L.S., A.J.v.Z. and M.H. managed the data. D.R. and J.E.K. provided resources. B.L.S., M.G.S. and H.N. wrote the manuscript, with contributions of all co-authors.

Acknowledgements

Funding

Conflict of interest
None declared

References

1. Sterken,M.G., Snoek,L.B., Kammenga,J.E. and Andersen,E.C. (2015) The laboratory domestication of *Caenorhabditis elegans*. *Trends Genet.*, **31**, 224–231.
2. Noble,L.M., Chang,A.S., McNelis,D. *et al.* (2015) Natural variation in plep-1 causes male-male Copulatory behavior in *C. elegans*. *Curr. Biol.*, **25**, 2730–2737.
3. Andersen,E.C., Bloom,J.S., Gerke,J.P. and Kruglyak,L. (2014) A variant in the neuropeptide receptor npr-1 is a major determinant of *Caenorhabditis elegans* growth and physiology. *PLoS Genet.*, **10**, e1004156.
4. Ghosh,R., Andersen,E.C., Shapiro,J.A. *et al.* (2012) Natural variation in a chloride channel subunit confers avermectin resistance in *C. elegans*. *Science*, **335**, 574–578.
5. Kammenga,J.E., Doroszuk,A., Riksen,J.A. *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.
6. Large,E.E., Xu,W., Zhao,Y. *et al.* (2016) Selection on a subunit of the NURF chromatin remodeler modifies life history traits in a domesticated strain of *Caenorhabditis elegans*. *PLoS Genet.*, **12**, e1006219.
7. Bendesky,A., Pitts,J., Rockman,M.V. *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.
8. Bendesky,A., Tsunozaki,M., Rockman,M.V. *et al.* (2011) Catecholamine receptor polymorphisms affect decision-making in *C. elegans*. *Nature*, **472**, 313–318.
9. O'Donnell,M.P., Chao,P.H., Kammenga,J.E. and Sengupta,P. (2018) Rictor/TORC2 mediates gut-to-brain signaling in the regulation of phenotypic plasticity in *C. elegans*. *PLoS Genet.*, **14**, e1007213.
10. Cook,D.E., Zdraljevic,S., Tanny,R.E. *et al.* (2016) The genetic basis of natural variation in *Caenorhabditis elegans* telomere length. *Genetics*, **204**, 371–383.
11. Schmid,T., Snoek,L.B., Frohli,E. *et al.* (2015) Systemic regulation of RAS/MAPK Signaling by the serotonin metabolite 5-HIAA. *PLoS Genet.*, **11**, e1005236.
12. Seidel,H.S., Ailion,M., Li,J. *et al.* (2011) A novel sperm-delivered toxin causes late-stage embryo lethality and transmission ratio distortion in *C. elegans*. *PLoS Biol.*, **9**, e1001115.
13. McGrath,P.T., Rockman,M.V., Zimmer,M. *et al.* (2009) Quantitative mapping of a digenic behavioral trait implicates globin variation in *C. elegans* sensory behaviors. *Neuron*, **61**, 692–699.
14. Reddy,K.C., Andersen,E.C., Kruglyak,L. and Kim,D.H. (2009) A polymorphism in npr-1 is a behavioral determinant of pathogen susceptibility in *C. elegans*. *Science*, **323**, 382–384.
15. Rogers,C., Persson,A., Cheung,B. and de Bono,M. (2006) Behavioral motifs and neural pathways coordinating O_2 responses and aggregation in *C. elegans*. *Curr. Biol.*, **16**, 649–659.
16. Gloria-Soria,A. and Azevedo,R.B. (2008) Npr-1 regulates foraging and dispersal strategies in *Caenorhabditis elegans*. *Curr. Biol.*, **18**, 1694–1699.

17. Seidel,H.S., Rockman,M.V. and Kruglyak,L. (2008) Widespread genetic incompatibility in *C. elegans* maintained by balancing selection. *Science*, **319**, 589–594.

18. Tijsterman,M., Okihara,K.L., Thijssen,K. and Plasterk,R.H. (2002) PPW-1, a PAZ/PIWI protein required for efficient germline RNAi, is defective in a natural isolate of *C. elegans*. *Curr. Biol.*, **12**, 1535–1540.

19. Palopoli,M.F., Rockman,M.V., TinMaung,A. *et al.* (2008) Molecular basis of the copulatory plug polymorphism in *Caenorhabditis elegans*. *Nature*, **454**, 1019–1022.

20. Reiner,D.J., Ailion,M., Thomas,J.H. and Meyer,B.J. (2008) *C. elegans* anaplastic lymphoma kinase ortholog SCD-2 controls dauer formation by modulating TGF-beta signaling. *Curr. Biol.*, **18**, 1101–1109.

21. Zdraljevic,S., Fox,B.W., Strand,C. *et al.* (2019) Natural variation in *C. elegans* arsenic toxicity is explained by differences in branched chain amino acid metabolism. *elife*, 8.

22. Hahnel,S.R., Zdraljevic,S., Rodriguez,B.C. *et al.* (2018) Extreme allelic heterogeneity at a *Caenorhabditis elegans* beta-tubulin locus explains natural resistance to benzimidazoles. *PLoS Pathog.*, **14**, e1007226.

23. Zdraljevic,S., Strand,C., Seidel,H.S. *et al.* (2017) Natural variation in a single amino acid substitution underlies physiological responses to topoisomerase II poisons. *PLoS Genet.*, **13**, e1006891.

24. Ben-David,E., Burga,A. and Kruglyak,L. (2017) A maternal-effect selfish genetic element in *Caenorhabditis elegans*. *Science*, **356**, 1051–1055.

25. Greene,J.S., Brown,M., Dobosiewicz,M. *et al.* (2016) Balancing selection shapes density-dependent foraging behaviour. *Nature*, **539**, 254–258.

26. Brady,S.C., Zdraljevic,S., Bisaga,K.W. *et al.* (2019) A novel gene underlies Bleomycin-response variation in *Caenorhabditis elegans*. *Genetics*., vol. **212** no. 4 page 1453–1468.

27. Thompson,O.A., Snoek,L.B., Nijveen,H. *et al.* (2015) Remarkably divergent regions punctuate the genome assembly of the *Caenorhabditis elegans* Hawaiian strain CB4856. *Genetics*, **200**, 975–989.

28. Li,Y., Alvarez,O.A., Gutteling,E.W. *et al.* (2006) Mapping determinants of gene expression plasticity by genetical genomics in *C. elegans*. *PLoS Genet.*, **2**, e222.

29. Rockman,M.V. and Kruglyak,L. (2006) Genetics of global gene expression. *Nat. Rev. Genet.*, **7**, 862–872.

30. Doroszuk,A., Snoek,L.B., Fradin,E. *et al.* (2009) A genome-wide library of CB4856/N2 introgression lines of *Caenorhabditis elegans*. *Nucleic Acids Res.*, **37**, e110.

31. Sterken,M.G., Bevers,R.P.J., Volkers,R.J.M. *et al.* (2019) Dissecting the eQTL micro-architecture in *Caenorhabditis elegans*. *BioRxiv*., https://doi.org/10.1101/651885.

32. Andersen,E.C., Gerke,J.P., Shapiro,J.A. *et al.* (2012) Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. *Nat. Genet.*, **44**, 285–290.

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

34. Snoek,B.L., Volkers,R.J.M., Nijveen,H. *et al.* (2019) A multi-parent recombinant inbred line population of *C. elegans* allows identification of novel QTLs for complex life history traits. *BMC Biol.*, **17**, 24.

35. 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 (Camb)*, **92**, 331–348.

36. Frezal,L. and Felix,M.A. (2015) *C. elegans* outside the petri dish. *elife*, **4**, e05849.

37. Schulenburg,H. and Felix,M.A. (2017) The natural biotic environment of *Caenorhabditis elegans*. *Genetics*, **206**, 55–86.

38. Volkers,R.J., Snoek,L.B., Hubar,C.J. *et al.* (2013) Gene-environment and protein-degradation signatures characterize genomic and phenotypic diversity in wild *Caenorhabditis elegans* populations. *BMC Biol.*, **11**, 93.

39. Laricchia,K.M., Zdraljevic,S., Cook,D.E. and Andersen,E.C. (2017) Natural variation in the distribution and abundance of transposable elements across the *Caenorhabditis elegans* species. *Mol. Biol. Evol.*.

40. Evans,K.S., Zhao,Y., Brady,S.C. *et al.* (2017) Correlations of genotype with climate parameters suggest *Caenorhabditis elegans* niche adaptations. *G3 (Bethesda)*, **7**, 289–298.

41. Balla,K.M., Andersen,E.C., Kruglyak,L. and Troemel,E.R. (2015) A wild *C. elegans* strain has enhanced epithelial immunity to a natural microsporidian parasite. *PLoS Pathog.*, **11**, e1004583.

42. Dirksen,P., Marsh,S.A., Braker,I. *et al.* (2016) The native microbiome of the nematode *Caenorhabditis elegans*: gateway to a new host-microbiome model. *BMC Biol.*, **14**, 38.

43. Samuel,B.S., Rowedder,H., Braendle,C. *et al.* (2016) *Caenorhabditis elegans* responses to bacteria from its natural habitats. *Proc. Natl. Acad. Sci. USA*, **113**, E3941–E3949.

44. Barriere,A. and Felix,M.A. (2014) Isolation of *C. elegans* and related nematodes. *WormBook*, 1–19.

45. Felix,M.A. and Braendle,C. (2010) The natural history of *Caenorhabditis elegans*. *Curr. Biol.*, **20**, R965–R969.

46. Wang,Y.A., Snoek,B.L., Sterken,M.G. *et al.* (2019) Genetic background modifies phenotypic and transcriptional responses in a *C. elegans* model of alpha-synuclein toxicity. *BMC Genomics*, **20**, 232.

47. 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.

48. Paaby,A.B., White,A.G., Riccardi,D.D. *et al.* (2015) Wild worm embryogenesis harbors ubiquitous polygenic modifier variation. *elife*, **4**.

49. Elvin,M., Snoek,L.B., Frejno,M. *et al.* (2011) A fitness assay for comparing RNAi effects across multiple *C. elegans* genotypes. *BMC Genomics*, **12**, 510.

50. Sterken,M.G., van Bemmelen van der Plaat,L., Riksen,J.A.G. *et al.* (2017) Ras/MAPK Modifier Loci Revealed by eQTL in Caenorhabditis elegans. *G3*, **7**, 3185–3193.

51. Kammenga,J.E. (2017) The background puzzle: how identical mutations in the same gene lead to different disease symptoms. *FEBS J.*, **284**, 3362–3373.

52. Grishkevich,V., Ben-Elazar,S., Hashimshony,T. *et al.* (2012) A genomic bias for genotype-environment interactions in *C. elegans*. *Mol. Syst. Biol.*, **8**, 587.

53. Nakad,R., Snoek,L.B., Yang,W. *et al.* (2016) Contrasting invertebrate immune defense behaviors caused by a single gene, the *Caenorhabditis elegans* neuropeptide receptor gene npr-1. *BMC Genomics*, **17**, 280.

54. Stastna,J.J., Snoek,L.B., Kammenga,J.E. and Harvey,S.C. (2015) Genotype-dependent lifespan effects in peptone deprived *Caenorhabditis elegans*. *Sci. Rep.*, **5**, 16259.

55. Snoek,L.B., Orbidans,H.E., Stastna,J.J. *et al.* (2014) Widespread genomic incompatibilities in Caenorhabditis elegans. *G3*, **4**, 1813–1823.

56. Green,J.W., Snoek,L.B., Kammenga,J.E. *et al.* (2013) Genetic mapping of variation in dauer larvae development in growing populations of *Caenorhabditis elegans*. *Heredity*, **111**, 306–313.

57. Rodriguez,M., Snoek,L.B., Riksen,J.A. *et al.* (2012) Genetic variation for stress-response hormesis in *C. elegans* lifespan. *Exp. Gerontol.*, **47**, 581–587.

58. Gutteling,E.W., Doroszuk,A., Riksen,J.A. *et al.* (2007) Environmental influence on the genetic correlations between life-history traits in *Caenorhabditis elegans*. *Heredity (Edinb)*, **98**, 206–213.

59. Gutteling,E.W., Riksen,J.A., Bakker,J. and Kammenga,J.E. (2007) Mapping phenotypic plasticity and genotype-environment interactions affecting life-history traits in *Caenorhabditis elegans*. *Heredity (Edinb)*, **98**, 28–37.

60. Capra,E.J., Skrovanek,S.M. and Kruglyak,L. (2008) Comparative developmental expression profiling of two *C. elegans* isolates. *PLoS One*, **3**, e4055.

61. Viñuela,A., Snoek,L.B., Riksen,J.A. *et al.* (2012) Aging uncouples heritability and expression-QTL in *Caenorhabditis elegans*. *G3 (Bethesda)*, **2**, 597–605.

62. Jansen,R.C. (2003) Studying complex biological systems using multifactorial perturbation. *Nat. Rev. Genet.*, **4**, 145–151.

63. Jansen,R.C. and Nap,J.P. (2001) Genetical genomics: the added value from segregation. *Trends Genet.*, **17**, 388–391.

64. Viñuela,A., Snoek,L.B., Riksen,J.A. *et al.* (2010) Genome-wide gene expression regulation as a function of genotype and age in *C. elegans*. *Genome Res.*, **20**, 929–937.

65. Snoek,B.L., Sterken,M.G., Bevers,R.P.J. *et al.* (2017) Contribution of trans regulatory eQTL to cryptic genetic variation in *C. elegans*. *BMC Genomics*, **18**, 500.

66. Li,Y., Breitling,R., Snoek,L.B. *et al.* (2010) Global genetic robustness of the alternative splicing machinery in *Caenorhabditis elegans*. *Genetics*, **186**, 405–410.

67. Rockman,M.V., Skrovanek,S.S. and Kruglyak,L. (2010) Selection at linked sites shapes heritable phenotypic variation in *C. elegans*. *Science*, **330**, 372–376.

68. Terpstra,I.R., Snoek,L.B., Keurentjes,J.J. *et al.* (2010) Regulatory network identification by genetical genomics: signaling downstream of the Arabidopsis receptor-like kinase ERECTA. *Plant Physiol.*, **154**, 1067–1078.

69. Keurentjes,J.J., Fu,J., Terpstra,I.R. *et al.* (2007) Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. *Proc. Natl. Acad. Sci. USA*, **104**, 1708–1713.

70. Zhu, M.M. and Wu, Q. (2008) Transcription network construction for large-scale microarray datasets using a high-performance computing approach. *BMC Genomics*, **9** Suppl 1, S5.

71. Vignes,M., Vandel,J., Allouche,D. *et al.* (2011) Gene regulatory network reconstruction using Bayesian networks, the Dantzig selector, the lasso and their meta-analysis. *PLoS One*, **6**, e29165.

72. Valba,O.V., Nechaev,S.K., Sterken,M.G. *et al.* (2015) On predicting regulatory genes by analysis of functional networks in *C. elegans*. *BioData Min*, **8**, 33.

73. Kliebenstein,D.J., West,M.A., van Leeuwen,H. *et al.* (2006) Identification of QTLs controlling gene expression networks defined a priori. *BMC Bioinform*, **7**, 308.

74. Serin,E.A., Nijveen,H., Hilhorst,H.W. and Ligterink,W. (2016) Learning from co-expression networks: possibilities and challenges. *Front. Plant Sci.*, **7**, 444.

75. Snoek,L.B., Terpstra,I.R., Dekter,R. *et al.* (2012) Genetical genomics reveals Large scale genotype-by-environment interactions in Arabidopsis thaliana. *Front. Genet.*, **3**, 317.

76. Nijveen,H., Ligterink,W., Keurentjes,J.J. *et al.* (2017) AraQTL - workbench and archive for systems genetics in Arabidopsis thaliana. *Plant J.*, **89**, 1225–1235.

77. van der Velde,K.J., de Haan,M., Zych,K. *et al.* (2014) WormQTLHD–a web database for linking human disease to natural variation data in *C. elegans*. *Nucleic Acids Res.*, **42**, D794–D801.

78. Snoek,L.B., Joeri van der Velde,K., Li,Y. *et al.* (2014) Worm variation made accessible: take your shopping cart to store, link, and investigate. *Worm*, **3**, e28357.

79. Snoek,L.B., Van der Velde,K.J., Arends,D. *et al.* (2013) WormQTL—public archive and analysis web portal for natural variation data in Caenorhabditis spp. *Nucleic Acids Res.*, **41**, D738–D743.

80. Singh,K.D., Roschitzki,B., Snoek,L.B. *et al.* (2016) Natural genetic variation influences protein abundances in *C. elegans* developmental signalling pathways. *PLoS One*, **11**, e0149418.

81. Gao,A.W., Sterken,M.G., Uit de,J. *et al.* (2018) Natural genetic variation in *C. elegans* identified genomic loci controlling metabolite levels. *Genome Res.*, **28**, 1296–1308.

82. Andersen,E.C., Shimko,T.C., Crissman,J.R. *et al.* (2015) A powerful new quantitative genetics platform, combining *Caenorhabditis elegans* high-throughput fitness assays with a Large collection of recombinant strains. *G3 (Bethesda)*, **5**, 911–920.

83. Evans,K.S., Brady,S.C., Bloom,J.S. *et al.* (2018) Shared genomic regions underlie natural variation in diverse toxin responses. *Genetics*, **210**, 1509–1525.

84. Gaertner,B.E., Parmenter,M.D., Rockman,M.V. *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.

85. Glater,E.E., Rockman,M.V. and Bargmann,C.I. (2014) Multigenic natural variation underlies *Caenorhabditis elegans* olfactory preference for the bacterial pathogen Serratia marcescens. *G3 (Bethesda)*, **4**, 265–276.

86. Greene,J.S., Dobosiewicz,M., Butcher,R.A. *et al.* (2016) Regulatory changes in two chemoreceptor genes contribute to a *Caenorhabditis elegans* QTL for foraging behavior. *elife*, **5**, pii: e21454.

87. Harvey,S.C. (2009) Non-dauer larval dispersal in *Caenorhabditis elegans*. *J. Exp. Zool. B Mol. Dev. Evol.*, **312B**, 224–230.

88. Harvey,S.C., Shorto,A. and Viney,M.E. (2008) Quantitative genetic analysis of life-history traits of *Caenorhabditis elegans* in stressful environments. *BMC Evol. Biol.*, **8**, 15.

89. Lee,D., Yang,H., Kim,J. *et al.* (2017) The genetic basis of natural variation in a phoretic behavior. *Nat. Commun.*, **8**, 273.

90. Zhu,Z., Lu,Q., Zeng,F. *et al.* (2015) Compatibility between mitochondrial and nuclear genomes correlates with the quantitative trait of lifespan in *Caenorhabditis elegans*. *Sci. Rep.*, **5**, 17303.

91. Rockman,M.V. and Kruglyak,L. (2009) Recombinational landscape and population genomics of *Caenorhabditis elegans*. *PLoS Genet.*, **5**, e1000419.

92. Rockman,M.V. and Kruglyak,L. (2008) Breeding designs for recombinant inbred advanced intercross lines. *Genetics*, **179**, 1069–1078.

93. Kamkina,P., Snoek,L.B., Grossmann,J. *et al.* (2016) Natural genetic variation differentially affects the proteome and Transcriptome in *Caenorhabditis elegans*. *Mol. Cell. Proteomics*, **15**, 1670–1680.

94. Hibbs,M.A., Hess,D.C., Myers,C.L. *et al.* (2007) Exploring the functional landscape of gene expression: directed search of large microarray compendia. *Bioinformatics*, **23**, 2692–2699.

95. Sinha,A., Rae,R., Iatsenko,I. and Sommer,R.J. (2012) System wide analysis of the evolution of innate immunity in the nematode model species *Caenorhabditis elegans* and Pristionchus pacificus. *PLoS One*, **7**, e44255.

96. Turek,M., Besseling,J., Spies,J.P. *et al.* (2016) Sleep-active neuron specification and sleep induction require FLP-11 neuropeptides to systemically induce sleep. *elife*, **5**.

97. Cho,A., Shin,J., Hwang,S. *et al.* (2014) WormNet v3: a network-assisted hypothesis-generating server for *Caenorhabditis elegans*. *Nucleic Acids Res.*, **42**, W76–W82.

98. Grove,C., Cain,S., Chen,W.J. *et al.* (2018) Using WormBase: a genome biology resource for *Caenorhabditis elegans* and related nematodes. *Methods Mol. Biol.*, **1757**, 399–470.

99. Brown,J.B. and Celniker,S.E. (2015) Lessons from modENCODE. *Annu. Rev. Genomics Hum. Genet.*, **16**, 31–53.

100. Kudron,M.M., Victorsen,A., Gevirtzman,L. *et al.* (2018) The ModERN resource: genome-wide binding profiles for hundreds of drosophila and *Caenorhabditis elegans* transcription factors. *Genetics*, **208**, 937–949.

101. Warde-Farley,D., Donaldson,S.L., Comes,O. *et al.* (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.*, **38**, W214–W220.

102. Szklarczyk,D., Franceschini,A., Wyder,S. *et al.* (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.*, **43**, D447–D452.

103. Glauser,D.A., Chen,W.C., Agin,R. *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.

104. Alberts,R., Terpstra,P., Li,Y. *et al.* (2007) Sequence polymorphisms cause many false cis eQTLs. *PLoS One*, **2**, e622.

105. Zych,K., Snoek,B.L., Elvin,M. *et al.* (2017) reGenotyper: detecting mislabeled samples in genetic data. *PLoS One*, **12**, e0171324.

106. West,M.A., van Leeuwen,H., Kozik,A. *et al.* (2006) High-density haplotyping with microarray-based expression and single feature polymorphism markers in Arabidopsis. *Genome Res.*, **16**, 787–795.

107. West,M.A., Kim,K., Kliebenstein,D.J. *et al.* (2007) Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in Arabidopsis. *Genetics*, **175**, 1441–1450.

108. Smith,E.N. and Kruglyak,L. (2008) Gene-environment interaction in yeast gene expression. *PLoS Biol.*, **6**, e83.

109. Cubillos,F.A., Stegle,O., Grondin,C. *et al.* (2014) Extensive cis-regulatory variation robust to environmental perturbation in Arabidopsis. *Plant Cell*, **26**, 4298–4310.

110. Zych,K., Li,Y., van der Velde,J.K. *et al.* (2015) Pheno2Geno—high-throughput generation of genetic markers and maps from molecular phenotypes for crosses between inbred strains. *BMC Bioinform*, **16**, 51.

111. Noble,L.M., Chelo,I., Guzella,T. *et al.* (2017) Polygenicity and epistasis underlie fitness-proximal traits in the *Caenorhabditis elegans* multiparental experimental evolution (CeMEE) panel. *Genetics*, **207**, 1663–1685.

112. Serin,E.A.R., Snoek,L.B., Nijveen,H. *et al.* (2017) Construction of a high-density genetic map from RNA-Seq data for an Arabidopsis bay-0 × Shahdara RIL population. *Front. Genet.*, **8**, 201.

113. Shimko,T.C. and Andersen,E.C. (2014) COPASutils: an R package for reading, processing, and visualizing data from COPAS large-particle flow cytometers. *PLoS One*, **9**, e111090.

114. Green,J.W., Stastna,J.J., Orbidans,H.E. and Harvey,S.C. (2014) Highly polygenic variation in environmental perception determines dauer larvae formation in growing populations of *Caenorhabditis elegans*. *PLoS One*, **9**, e112830.

115. Bernstein,M.R. and Rockman,M.V. (2016) Fine-scale crossover rate variation on the *Caenorhabditis elegans* X chromosome. *G3 (Bethesda)*, **6**, 1767–1776.

116. Kawakatsu,T., Huang,S.S., Jupe,F. *et al.* (2016) Epigenomic diversity in a global collection of Arabidopsis thaliana accessions. *Cell*, **166**, 492–505.

117. Van Nostrand,E.L. and Kim,S.K. (2013) Integrative analysis of *C. elegans* modENCODE ChIP-seq data sets to infer gene regulatory interactions. *Genome Res.*, **23**, 941–953.

118. Contrino,S., Smith,R.N., Butano,D. *et al.* (2012) modMine: flexible access to modENCODE data. *Nucleic Acids Res.*, **40**, D1082–D1088.

119. Altschul,S.F., Gish,W., Miller,W. *et al.* (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.

120. Smyth,G.K. and Speed,T. (2003) Normalization of cDNA microarray data. *Methods*, **31**, 265–273.

121. Ritchie,M.E., Phipson,B., Wu,D. *et al.* (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, **43**, e47.

122. Zahurak,M., Parmigiani,G., Yu,W. *et al.* (2007) Pre-processing Agilent microarray data. *BMC Bioinform*, **8**, 142.

123. Smedley,D., Haider,S., Durinck,S. *et al.* (2015) The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res.*, **43**, W589–W598.

124. Broman,K.W. (2015) R/qtlcharts: interactive graphics for quantitative trait locus mapping. *Genetics*, **199**, 359–361.

125. van der Bent,M.L., Sterken,M.G., Volkers,R.J. *et al.* (2014) Loss-of-function of beta-catenin bar-1 slows development and activates the Wnt pathway in *Caenorhabditis elegans*. *Sci. Rep.*, **4**, 4926.

126. Snoek,L.B., Sterken,M.G., Volkers,R.J. *et al.* (2014) A rapid and massive gene expression shift marking adolescent transition in *C. elegans*. *Sci. Rep.*, **4**, 3912.