frontiers in Genetics

ORIGINAL RESEARCH
published: 03 November 2020
doi: 10.3389/fgene.2020.501376

Dissecting the eQTL Micro-Architecture in *Caenorhabditis elegans*

Mark G. Sterken[1], Roel P. J. Bevers[1], Rita J. M. Volkers[1], Joost A. G. Riksen[1], Jan E. Kammenga[1] and Basten L. Snoek[1,2]**

[1] Laboratory of Nematology, Wageningen University & Research, Wageningen, Netherlands, [2] Theoretical Biology & Bioinformatics, Utrecht University, Utrecht, Netherlands

The study of expression quantitative trait loci (eQTL) using natural variation in inbred populations has yielded detailed information about the transcriptional regulation of complex traits. Studies on eQTL using recombinant inbred lines (RILs) led to insights on *cis* and *trans* regulatory loci of transcript abundance. However, determining the underlying causal polymorphic genes or variants is difficult, but ultimately essential for the understanding of regulatory networks of complex traits. This requires insight into whether associated loci are single eQTL or a combination of closely linked eQTL, and how this QTL micro-architecture depends on the environment. We addressed these questions by testing for independent replication of previously mapped eQTL in *Caenorhabditis elegans* using new data from introgression lines (ILs). Both populations indicate that the overall heritability of gene expression, number, and position of eQTL differed among environments. Across environments we were able to replicate 70% of the *cis*- and 40% of the *trans*-eQTL using the ILs. Testing eight different simulation models, we suggest that additive effects explain up to 60–93% of RIL/IL heritability for all three environments. Closely linked eQTL explained up to 40% of RIL/IL heritability in the control environment whereas only 7% in the heat-stress and recovery environments. In conclusion, we show that reproducibility of eQTL was higher for *cis* vs. *trans* eQTL and that the environment affects the eQTL micro-architecture.

Keywords: genetic architecture, eQTL, *Caenorhabditis elegans*, introgression line, recombinant inbred line

OPEN ACCESS

Edited by:
Natalia Polouliakh, Sony Computer Science Laboratories, Inc., Japan

Reviewed by:
Nathan Weinstein, Universidad Nacional Autónoma de México, Mexico
Wankui Gong, Institute of Cotton Research (CAAS), China

****Correspondence:***
Mark G. Sterken
mark.sterken@wur.nl
Basten L. Snoek
l.b.snoek@uu.nl

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Genetics

Received: *29 September 2019*
Accepted: *13 October 2020*
Published: *03 November 2020*

Citation:
Sterken MG, Bevers RPJ, Volkers RJM, Riksen JAG, Kammenga JE and Snoek BL (2020) Dissecting the eQTL Micro-Architecture in Caenorhabditis elegans. Front. Genet. 11:501376. doi: 10.3389/fgene.2020.501376

INTRODUCTION

The genetic architecture of quantitative traits in genetically segregated populations such as recombinant inbred lines (RILs) differs within and across species and strongly depends on the environment and type of trait (Fournier and Schacherer, 2017). In some cases, quantitative traits are (nearly) monogenic and display Mendelian characteristics. For instance the Kallman syndrome in humans is caused by a knockout of FGFR1 (which encodes fibroblast growth factor receptor 1) (Muenke et al., 1994), industrial melanism in British peppered moths is caused by a single gene mutation (van't Hof et al., 2011), and natural variation in the *npr-1* gene in the nematode *Caenorhabditis elegans* is a major determinant of multiple life-history traits (Andersen et al., 2014; Sterken et al., 2015). These monogenic traits are relatively easy to detect but rather exceptions than the rule. Most quantitative traits display a complex polygenic genetic

architecture, regulated by multiple QTLs, where each locus explains a proportion of the trait variation (Albert and Kruglyak, 2015).

Studying quantitative genetics of the transcriptome has proven useful in understanding the architecture of complex traits. The transcriptional architecture can be defined as polymorphic regulators affecting the expression of genes throughout the genome. These polymorphic regulators can affect gene expression in direct ways [e.g., kinase cascades (Terpstra et al., 2010)], or indirectly [e.g., by affecting behavior (Andersen et al., 2014; Sterken et al., 2015)]. Expression quantitative trait loci (eQTL) can be measured reliably in a high-throughput fashion in many organisms like: yeast, *Arabidopsis thaliana*, tomato, maize, *C. elegans*, mice, and humans (Schadt et al., 2003; Brem and Kruglyak, 2005; Li et al., 2006; Keurentjes et al., 2007; Ranjan et al., 2016). The advantage of the eQTL-approach lies in the assessment of thousands of traits simultaneously for which the genetic architectures can be compared (Jansen and Nap, 2001; Gilad et al., 2008). First, the variance explained per underlying QTL will cover the whole range from essentially monogenic (Mendelian) to nearly undetectable. Second, correlating gene expression to loci can reveal the biological processes and phenotypes that lay downstream of (e)QTL (e.g., Jimenez-Gomez et al., 2010, 2011; Terpstra et al., 2010; Andersen et al., 2014; Schmid et al., 2015; Sterken et al., 2017; Albert et al., 2018). In general, two types of eQTL can be defined: (i) *trans*-eQTL where a single locus affects the expression of many genes across the genome and (ii) *cis*-eQTL where the locus is located on or near the gene of which it affects the expression (Nica and Dermitzakis, 2013; Albert et al., 2018). *Cis*-eQTL can occur due to polymorphisms affecting the transcription of the gene under study, e.g., deletions or non-functional promotor regions. eQTL hotspots are loci that regulate a high abundance of transcripts. These hotspots are also called *trans*-bands. The occurrence of *trans*-bands was found to be specific for environmental conditions linked to genetic effects. Together, the *cis*-eQTL and *trans*-bands typically cover ∼75% of the mapped eQTL in any study (e.g., Snoek et al., 2017; Albert et al., 2018).

The detection of (e)QTL depends on many different factors like developmental stage, background mutations, and the ambient environment, to name a few (Li et al., 2006; Vinuela et al., 2010; Duveau and Felix, 2012; Snoek et al., 2012; Andersen et al., 2014; Francesconi and Lehner, 2014). Importantly, QTL detection is mainly determined by the number of genetically segregating RILs and the density and structure of the genetic marker map. Further characterization of QTLs often involves the identification of the underlying causal polymorphic genes in order to understand molecular mechanisms affecting complex traits and/or to facilitate breeding for selected traits. An important hallmark of QTL mapping has been the fine mapping and identification of polymorphic regulators or quantitative trait nucleotides (QTN) (Radwan and Babik, 2012; Rockman, 2012).

Fine mapping of causal variants requires determining if associated loci are single QTL or combinations of closely linked QTL, here coined the QTL micro-architecture, and how this micro-architecture depends on the environment. Apparent single QTL may be truly single QTL or a combination of closely linked QTL not segregating in the studied population. These are often hard to characterize because of a lack of resolution of the genetic map or shortage of a sufficient number of recombination events. Here, we set out to investigate the QTL micro-architecture of eQTL in the nematode *C. elegans* in a segregated population of RILs derived from a cross between wild-types Bristol N2 and CB4856 (Li et al., 2006; Thompson et al., 2015). Briefly, Bristol N2 and CB4856 wild types were allowed to mate after which the recombinant offspring were selfed (*C. elegans* is a hermaphrodite but males do occur and can be used for making crosses between strains). This yields homozygous RILs that were analyzed for their transcriptomes (Li et al., 2006). Subsequent association of polymorphic SNPs with transcriptional variation yielded eQTL. We use a set of previously detected eQTL and study their replication in a separate, independent, population of introgression lines (ILs) in this study (Doroszuk et al., 2009; Thompson et al., 2015; Snoek et al., 2017). Compared to the RILs, which are genetic mosaics of loci derived from both parents, ILs contain a single genomic segment of one parent in a genetic background of another parent. Typically, introgression lines are used to verify the existence of QTL, or to narrow down the location of a QTL in order to find the underlying causal polymorphism (e.g., see Gao et al., 2018). The RIL and IL populations are exposed to three environments: heat-shock, recovery from heat-shock, and a control environment at a standard rearing temperature. We investigate three aspects of eQTL mapped in RILs: (i) we address how well eQTL are replicated in ILs, (ii) we test if the predicted numbers of eQTL agree with the observed number of differentially expressed genes in ILs, and (iii) we simulate QTL architectures explaining the relative heritability in RILs and ILs and match our data to these models. Finally, we show that the eQTL microarchitecture under ambient conditions mainly consists of closely linked eQTL.

MATERIALS AND METHODS

Strains Used

The wild-type strains N2 and CB4856 were used and 57 introgression lines (ILs) with segments of CB4856 in an N2 genetic background (Doroszuk et al., 2009). Most of the ILs have been sequenced, and the genetic map was based on the sequenced genotypes (Thompson et al., 2015), a file with the strains and the map has been included in **Supplementary Table 1**. The data used for the recombinant inbred lines (RILs) is accessible in (Snoek et al., 2017).

Nematode Culturing

The strains were cultured as described previously (Snoek et al., 2017). In short, strains were kept in maintenance at 12°C on 6-cm Nematode Growth Medium (NGM) Petri dishes seeded with the *Escherichia coli* strain OP50 as food source (Brenner, 1974). Before starting the experiment, single hermaphrodites of each strain in the L2 stage were sub-cultured in 12-wells plates and grown at 20°C. The offspring was screened for the occurrence of males by microscopy and only populations consisting solely of hermaphrodites were transferred to 9-cm

NGM Petri dishes containing *E. coli* OP50 and grown until the bacterial food was depleted.

Treatments for the Transcriptomic Experiment

The experiment described in our previous paper was repeated on the IL population and the parental strains (Snoek et al., 2017). We started the experiment by transferring a starved population to a new 9-cm NGM Petri dish seeded with *E. coli* OP50. After 60 h at 20°C, the populations consisting of egg-laying adults were bleached to obtain the eggs, which were transferred to a new 9-cm NGM Petri dish (Brenner, 1974). Around 46 h the developmental stage of the population was assessed by microscopy, verifying that the population consisted mainly of L4 larvae. Populations not consisting of L4 larvae were not used. The strains were exposed to one of three environments: (i) a control environment; grown for 48 h at 20°C, (ii) a heat-stress environment; grown for 46 h at 20°C and subsequently exposed to 35°C for 2 h, and (iii) a recovery environment; grown for 46 h at 20°C, exposed to 35°C for 2 h, and thereafter returned to 20°C for 2 h. Directly after the treatment the animals were washed off the Petri dish using M9 buffer and collected in a 1.5 mL Eppendorf tube, which was centrifuged and the pellet was flash frozen in liquid nitrogen.

RNA Isolation, cDNA Synthesis, Labeling and Hybridization

The procedure was followed as described before (Snoek et al., 2017). The RNA of the samples was isolated using the RNEasy Micro Kit from Qiagen (Hilden, Germany) following the 'Purification of Total RNA from Animal and Human Tissues.' The lysis step was modified, pellets were lysed in 150 µl RLT buffer, 295 µl RNAse-free water, 800 µg/ml proteinase K and 1% β-mercaptoethanol, which was incubated at 55°C and 1,000 rpm in a Thermomixer (Eppendorf, Hamburg, Germany) for 30 min, where after the manufacturer's protocol was followed.

Before starting cDNA synthesis, the quality and quantity of the RNA was measured by NanoDrop-1000 spectrophotometer (Thermo Scientific, Wilmington DE, United States). The integrity of the RNA was assessed by loading 3 µL per sample on a 1% agarose gel. Samples with good quality RNA were processed following the 'Two-Color Microarray-Based Gene Expression Analysis; Low Input Quick Amp Labeling' -protocol, version 6.0 from Agilent (Agilent Technologies, Santa Clara, CA, United States). The microarrays used were Agilent *C. elegans* (V2) Gene Expression Microarray 4 × 44K slides. After quality control of the RNA isolation, we obtained 56/57 ILs for the control and heat-stress environment, and 55/57 ILs for the recovery environment. The parental strains were ran in four replicates, except for the CB4856 recovery experiment, which was ran in five replicates. For the parental strain N2: four control samples, four heat stress samples and three recovery samples passed quality control. For the parental strain CB4856: three control samples, three heat stress samples and five recovery samples passed quality control.

Array Scanning, Data Extraction, and Normalization

After washing, the microarrays were scanned using an Agilent High Resolution C Scanner following the recommended settings. For data extraction, Agilent Feature Extraction Software (version 10.7.1.1) was used, following manufacturer's guidelines. The extracted data was normalized together with the previously published RIL data (Snoek et al., 2017) using the limma package in "R" (version 3.4.2, x64) (Ritchie et al., 2015). As recommended, the data was not background corrected before normalization (Zahurak et al., 2007), the within-array normalization used the 'Loess' method and between-array normalization used the Quantile method (Smyth and Speed, 2003).

Data Analysis

Data was analyzed using "R" (version 3.4.2, x64) with custom written scripts (R Core Team, 2017), accessible via https://git.wur.nl/published_papers/sterken_2019_closely_linked_qtl. For analysis, the dplyr and tidyr packages were used for data organization (Wickham, 2018; Wickham et al., 2018), and plots were generated using ggplot2 (Wickham, 2009), except for plots displaying simulation results. The transcriptome data analyzed was deposited at ArrayExpress under E-MTAB-5779 (RIL data) (Snoek et al., 2017), and E-MTAB-7424 (IL and parental data, described in this paper).

eQTL Data

The eQTL data from the RILs were obtained from a previous publication (Snoek et al., 2017) and can be explored at www.bioinformatics.nl/wormqtl2 (Snoek et al., 2020). In short, previously eQTL were mapped using a single marker model, which was conducted separately for each environment. The threshold for each of the three environments was set at $-\log_{10}(p) > 3.9$ (false discovery rate, FDR = 0.05) (see Snoek et al., 2017 for details). The *cis*-eQTL were defined as the gene being inside a 1 Mb window of the QTL peak or within the confidence interval of the QTL peak. The confidence interval was defined as a 1.5 drop in LOD-score from the peak.

Data Transformation

The expression data was transformed to a *z*-score based on the standard deviation and mean of the N2-replicates per environment, using

$$Z_{N2,i,j} = \frac{I_{i,j} - \mu_{i,N2}}{\sigma_{i,N2}}$$

where Z_{N2} is the *Z*-score of spot *i* (1, 2, …, 45220) of strain *j* (one of the RILs or ILs) and *I* is the normalized intensity. The µ is the mean of the intensity over the N2 strains for spot *i* and the σ is the standard deviation over the N2 strains for spot *i*. This transformation shows the effect of the CB4856 introgressions in the ILs. It should be noted that these values cannot be directly interpreted as significances.

Additionally, the expression data was transformed to a *z*-score based on the standard deviation and mean of the CB4856-replicates per environment, using

$$Z_{CB4856,\ i,j} = \frac{I_{i,j} - \mu_{i,CB4856}}{\sigma_{i,CB4856}}$$

where Z_{CB4856} is the *Z*-score of spot *i* (1, 2,. . ., 45220) of strain *j* (one of the RILs or ILs) and *I* is the normalized intensity. The μ is the mean of the intensity over the CB4856 strains for spot *i* and the σ is the standard deviation over the CB4856 strains for spot *i*. This transformation shows the effect of the genetic background in the ILs. As for Z_{N2}, it should be noted that these values cannot be directly interpreted as significances.

In order to compare IL gene expression with eQTL effect sizes, another data transformation was used, as there the number of standard deviations difference with the mean is less informative than the effect in $\log_2$-ratio with the mean. Hence, we transformed the data by

$$R_{N2,i,j} = \log_2 \frac{I_{i,j}}{\mu_{i,N2}}$$

where R_{N2} is the $\log_2$-ratio with the mean of spot *i* of strain *j*, *I* is the normalized intensity, and μ is the mean intensity over the N2 samples.

Confirmation of IL Genotype

Before mapping, we assured that samples were labeled correctly by applying a *cis*-eQTL based analysis of gene expression compared to the genetic map of the IL population, for details and explanation, see (Zych et al., 2017). Samples of three strains were removed (WN263, WN264, and WN284), as the strains did not match the genotypes in the IL population (Doroszuk et al., 2009; Thompson et al., 2015).

Confirmation of eQTL in Introgression Lines

The differential expression per IL was calculated based on the *Z*-score calculated using the N2 parental strains (Z_{N2}). As each IL was measured once, we estimated the significance assuming a normal distribution based on the genetic-background parent. These significances were corrected for multiple testing by the Benjamini–Hochberg method, as applied by the *p.adjust* function in R (Benjamini and Hochberg, 1995). As a control and benchmark, the same procedure was applied to the individual RILs. The number of differentially expressed genes with an expected eQTL per CB4856 locus were counted for both the ILs and the RILs. These were expressed as a percentage of the expected eQTL in that locus (separately for *cis*- and *trans*-eQTL).

Another method of confirming eQTL is by correlating the differential expression per inbred line compared to the genetic background-parent (R_{N2}) with the QTL effect. By calculating the Pearson correlation of R_{N2} with the eQTL effect per IL or RIL separately for the *cis*- and *trans*-eQTL, it can be assessed how well the overall QTL patterns were recapitulated per strain. These correlations were calculated using the *cor* function and the significances of correlation were calculated using the *cor.test* function.

Confirmation of *Trans*-bands Using ILs

In the eQTL study on the RILs (Snoek et al., 2017), 19 *trans*-bands were identified (**Supplementary Table 3**). For each *trans*-band 2 to 6 ILs covering or flanking the locus were tested by correlating the *trans*-eQTL effect with the R_{N2} in the individual ILs. Significance was determined based on the correlation values in the N2 samples. Only if the correlation was stronger than the highest-correlating N2 sample, it was called as significant. Only significant positive correlations (effect direction matches the eQTL model) were scored as confirming the *trans*-band.

To determine the overall false-positive rate of this analysis, we determined the correlation of the ILs that were not in or near the *trans*-band region. These correlations were compared to the correlation value at which the *trans*-band was called confirmed. This resulted in an overall false positive rate of 0.11. It should be noted that this approach assumes that *trans*-bands stem from a single location and that these locations were accurately pinpointed in the RIL mapping. Hence, the derived threshold can be considered strict.

Expectation of Differential Expressed Genes Based on eQTL

First, the number of differentially expressed genes in each RIL and IL was calculated using the contrast with either of the ancestral strains, Z_{N2} and Z_{CB4856}, at an estimated significance of $p < 1 \times 10^{-5}$ to apply a strickt threshold for declaring significance as the *Z*-scores were based on the variance of a few samples. Second, the expected number of differentially expressed genes was estimated for each RIL and IL based on the eQTL architecture in the RIL population [eQTL threshold, $-\log_{10}(p) > 3.9$; FDR = 0.05] (Snoek et al., 2017). These values were normalized to the RIL population (the RIL population average was set to 1), to compare the relative number of differentially expressed genes expected versus observed in the ILs.

Differential Gene-Expression in N2 and CB4856

Differential gene expression between the two parental strains was independently calculated for each environment using a linear model

$$\log_2(I_{i,j}) \sim G_j + e_i$$

where *I* is the normalized intensity of spot *i* (1, 2,. . ., 45220) of strain *j* (CB4956 or N2) is explained over genotype *G* and residual variance *e*. For the parental strain N2: $n = 4$ in control and heat-stress environments, $n = 3$ in recovery environment. For the parental strain CB4856: $n = 3$ in control and heat stress environments and $n = 5$ recovery in recovery environment. The obtained significances were corrected for multiple testing using the Benjamini–Hochberg method implemented in the *p.adjust* function in R (Benjamini and Hochberg, 1995). A threshold of FDR = 0.1 was taken as requirement for differential expression.

Estimation of Trait Genetic Architecture

Two approaches were taken to determine trait architectures using both RIL and IL population: heritability calculations and directly testing the amount of variance between both populations. Both approaches are based on the notion that the trait variance in the populations scale with trait complexity, where

$$\sigma^2_{trait} \propto \sigma^2_{QTL\ 1} + \sigma^2_{QTL\ 2} + \ldots + \sigma^2_{QTL\ n} + \sigma^2_e$$

here σ^2 is the trait variance, proportional to the variance captured by *n* QTL (which can be multiple loci interacting; epistasis) and the measurement error. The contrast between the RIL and the IL population is that the QTL segregating in the IL population will affect fewer strains per QTL, whereas in the RIL population each QTL will segregate on average in 50% of the strains. In other words, the IL population shares most QTL effects because the genetic background is the same, whereas in the RIL population most QTL segregate.

Heritability was calculated as in (Brem and Kruglyak, 2005; Keurentjes et al., 2007), where

$$H^2_{i,F} = \frac{\sigma^2_{i,F} - \sigma^2_{i,P}}{\sigma^2_{i,F}}$$

where H^2 was the heritability of spot *i* (1, 2,. . ., 45220) of population *F* (either RIL or IL) within one of the three environments. The σ^2 indicated the trait variance of spot *i*. Here the pooled trait variance of the parental lines N2 and CB4856 (denoted with *P*) was used as an estimate of the measurement error, which was subtracted from the trait variance of either the RIL or IL population. We used a permutation approach to determine the threshold for significant heritability (Brem and Kruglyak, 2005; Vinuela et al., 2012). Thereto, the gene expression values per spot per environment were randomized over the strains (RIL and IL separately, but together with the same parental lines) and recalculated the heritability as described above. This was repeated 1,000 times, and the 5% highest value for each spot was taken as the FDR = 0.05 threshold. Negative heritabilities (variance in the parental strains larger than in the RIL or IL populations were not taken along in the analysis).

The second measure was by determining the ratio of trait variance between the IL and the RIL population. The assumption was that the variance contributed by random sources (σ^2_e) was equal in both populations. Hence, this is a measure of relative heritability and was calculated as

$$R_{H,i} = \log_2\left(\frac{\sigma^2_{i,IL}}{\sigma^2_{i,RIL}}\right)$$

where R_H was the heritability ratio of spot *i* (1, 2,. . ., 45220), and σ^2 was the trait variance in either the IL or the RIL population within one of the tree environments. We took the $\log_2$ ratio, meaning that traits with a higher heritability in the IL population resulted in a positive number, and traits with a higher heritability in the RIL population with a negative number. We performed two tests to place the relative heritability in perspective. First, we took the spots showing significant heritability (FDR = 0.05). Second, to test whether variance was significantly different between the two populations, we used the non-parametric Fligner–Killeen test for homogeneity of variances, this we corrected for multiple testing using the Benjamini–Hochberg method (FDR = 0.05) (Conover et al., 1981; Benjamini and Hochberg, 1995).

Simulating Trait Architecture Impact on Relative Heritability

To determine how the trait variance behaves proportionally in each population, we simulated different trait architectures in both the IL and RIL population. First, all QTL effect sizes were simulated by random drawing from a standard normal distribution ($\mu = 0$, $\sigma = 1$). These QTL were simulated in trait architectures containing 1 to 100 QTL (or 1 to 100 clusters of closely linked QTL, each 33331 times). To recapitulate the observed QTL distribution in *C. elegans*, we distributed these QTL along the informative markers in our IL and RIL population. This assured an enrichment of QTL on the chromosome arms compared to the chromosome tips and centers (Rockman et al., 2010; Snoek et al., 2017). To approach the residual variance (σ^2_e) in our data, we added random variation based on a normal distribution (*rnorm*) such that the average heritability of each simulated trait was 0.85

The following eight architectures were simulated: (i) randomly distributed additive QTL, (ii) randomly distributed additive QTL that only show their effect in one genotype, (iii) clusters of closely linked QTL with random effects, (iv) clusters of closely linked QTL that only show their effect in one genotype, (v) clusters of closely linked QTL that are balanced within the ancestral genotypes, (vi) clusters of closely linked QTL that are balanced within the ancestral genotypes and only show their effect in one genotype, (vii) randomly distributed interacting QTL, (viii) clusters of closely linked interacting QTL. The number of closely linked QTL per cluster was simulated from two to five. The additive and closely-linked model were similar in essence, but differ in the architecture, where we modeled the closely-linked QTL such that these were not segregated in the population used. The simulation functions can be found at the git archive: https://git.wur.nl/published_papers/sterken_2019_closely_linked_qtl.

The distribution of traits over a balanced closely-linked architecture and additive architecture was done by measuring the number of heritable traits with a heritability ratio between −3 and −2 (peak) and between 0.5 and 1.5 (intersect). These intervals were chosen as they represent areas where the distributions of these models differ. The number of expected traits per interval were counted as the median of all simulated architectures (so 1 to 100 QTL, and 2 to 5 QTL per closely-linked locus) in order to get an estimation that is less dependent on the chosen simulation parameters. The predicted number of traits at the peak and the intersect for the additive model (a) and the balanced closely-linked model (b) were used to solve the true distribution based on the real number of heritable traits in these heritability ratio intervals (c). This was done by solving the equations

$$a_{peak} \times x_1 + b_{peak} \times x_2 = c_{peak} \text{ and}$$

$$a_{intercept} \times x_1 + b_{intercept} \times x_2 = c_{intercept}$$

for x_1 and x_2 using the *solve* function in *R*. This led to the estimated number of additive traits (x_1) and balanced closely linked traits (x_2).

RESULTS

Genome-Wide Confirmation of eQTL Effects and Location Using Introgression Lines

Analysis of differential expression showed that introgression lines (ILs) could confirm both *cis*- and *trans*-eQTL that were previously mapped in the recombinant inbred line (RIL) population (Snoek et al., 2017). A population of 48 RILs, 56 ILs (Doroszuk et al., 2009; **Supplementary Table 1**), and at least three replicates of the ancestral strains N2 and CB4856 were exposed to the three environments (**Figure 1A**). Based on the number of eQTL expected in the CB4856 locus of each IL, we found that on average 69.5% (FDR = 0.05) of the *cis*-eQTL found using the RILs could be confirmed using the ILs over all three environments (FDR = 0.05). The *trans*-eQTL found in the RILs, were confirmed on average 42.2% (FDR = 0.05) in the ILs (**Figure 1** and **Supplementary Table 2**). Note that the tiling structure of the ILs meant that each QTL location was covered by 0–5 (5–95% quantiles) ILs (**Supplementary Figure 1**). The percentage of confirmed eQTL was positively correlated with the RIL significance threshold (**Supplementary Figure 2A**). Furthermore, correlation analysis of the $\log_2$ ratio with the N2 ancestor in the ILs versus the eQTL effect measured by RILs showed that eQTL effects in the CB4856 locus of each strain were well approximated in the ILs (average Pearson correlation, rho = 0.661; **Supplementary Figure 2B** and **Supplementary Table 2**). Together, over all three environments, the analyses confirmed eQTL occurrence and effect sizes in the RIL population in independent experiments using an IL population.

Next, we investigated the *trans*-bands identified in the RIL population more closely. The *trans*-bands were described in a previous study (Snoek et al., 2017). For each *trans*-band, we took 2 to 6 ILs covering and flanking the locus and tested if the *trans*-eQTL effects correlated significantly with differentially expressed genes in the ILs compared to N2. We found supporting evidence for 14 out of 19 *trans*-bands at an overall false-positive rate of 0.11; see Section "Materials and Methods." Furthermore, we narrowed down the location for some of these *trans*-bands (**Supplementary File 1** and **Supplementary Table 3**). In particular the *trans*-bands in the control and heat-stress environment were well supported (5/5 and 6/7, respectively), whereas those in the recovery environment were less (3/7). The poor support in the recovery environment is likely due to the few *trans*-eQTL per *trans*-band in that environment (median of 25 genes in recovery versus 51 in control and 125 in heat stress). Altogether, we conclude that *trans*-bands were highly replicable, to a higher extent than individual *trans*-eQTL.

Next, we asked if the number of differentially expressed genes in the IL population could be predicted from the RIL population. Therefore we compared the gene expression in both populations to the CB4856 ancestor to measure the effect of the N2 loci and to the N2 ancestor to measure the effect of the CB4856 loci. The effect of CB4856 introgressions in an N2 genetic background on gene expression was negatively correlated with the expectations based on the RIL population. In other words, in the RIL population the number of genes differentially expressed due to the introgression was under-estimated and the number of genes differentially expressed due to the genetic background was over-estimated (**Figure 2**). In the ILs we found that the relative number of differentially expressed genes compared to N2 was 6.7-fold higher than expected based on the differentially expressed genes found in the RILs (paired *t*-test, $p < 1 \times 10^{-54}$). Furthermore, when compared to CB4856, the number of differentially expressed genes in the ILs was 3.3-fold lower than expected from the RILs (paired *t*-test, $p < 1 \times 10^{-128}$). The differences between RILs and ILs were consistent when changing both the threshold for differential expression and the threshold for eQTL-based expectations (**Supplementary Figure 3**). These results show that the expectations from a single marker model are a poor predictor for differentially expressed genes in ILs. Importantly, a single marker model under-estimates the impact of small introgressions and over-estimates the impact of the genetic background.

Micro-Architecture of eQTL

The distribution and differences of genetic recombination over ILs and RILs allowed studying the architecture of eQTL in *C. elegans*. We hypothesized that different underlying trait architectures would leave different trait-variance distributions in the IL and RIL population. For example, a simple trait architecture dominated by one major QTL would only lead to an effect on trait levels in a single or a few ILs, but in ∼50% of the RILs. Hence, resulting in a low trait variance over the genome-wide IL population compared to the RIL population. As heritability was a straightforward way to interpret trait variance, we estimated the broad-sense heritability (H^2) in both populations (**Supplementary Table 4**). In general, heritability was higher in the RIL population. In total, over all conditions, we found 10673 unique genes that showed significant heritability in the RIL population, whereas we found 2,952 unique genes that showed significant heritability in the IL population (permutation, FDR = 0.05; **Supplementary Figure 4A**). Only 929 genes in the RILs and 8 genes in the ILs showed heritable gene expression variation over all three environments (**Supplementary Figures 4B,C**). The reason for the difference between populations was partly due to 534 out of 929 genes having an eQTL in the RILs (of which 354 had a *cis*-eQTL), which captured little heritable variation in the whole IL population. From the heritability analysis, we conclude that, as expected, heritability was higher in the RIL population, and that heritability is highly environment dependent.

To gain insight in the impact of different trait architectures, we simulated eight distinct architectures, including additive and epistatic architectures (see section "Materials and Methods," **Supplementary Figures 5, 6**) and measured trait variance ratios between the IL and RIL population. Subsequently, we

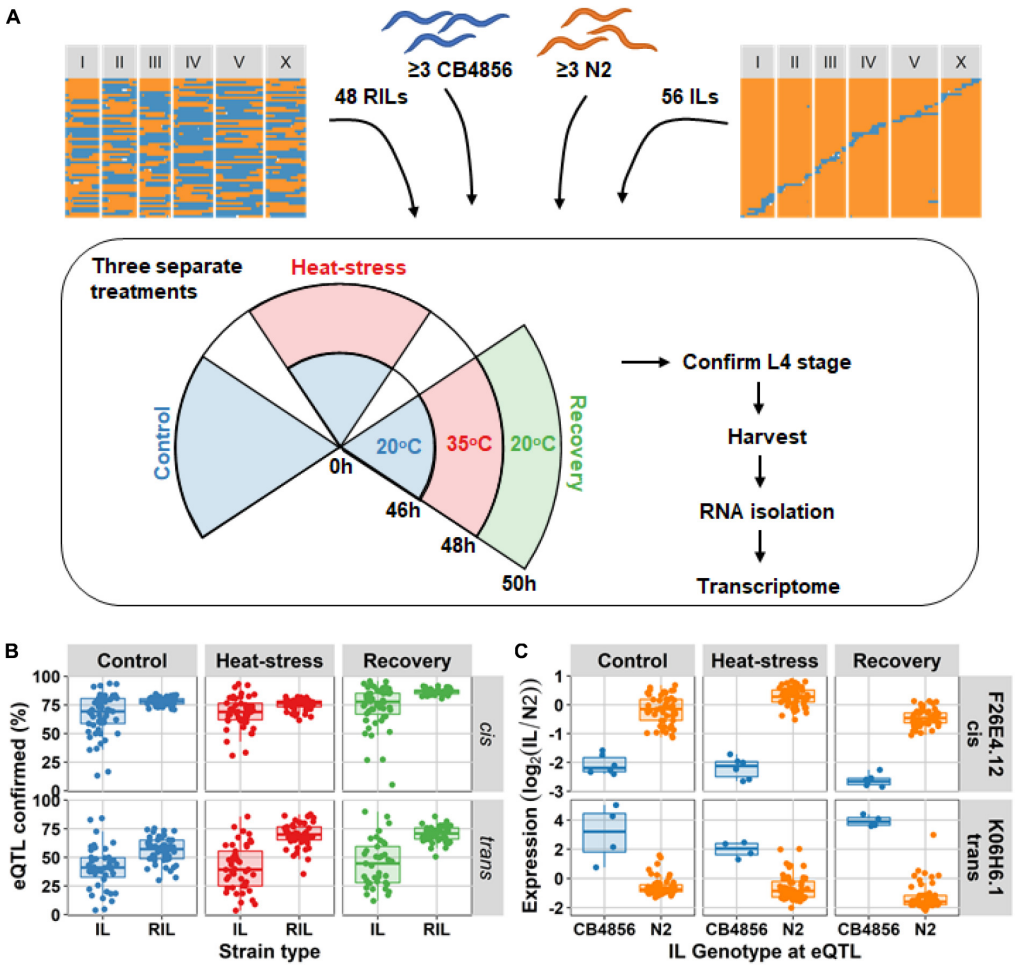

FIGURE 1 | Experimental setup and eQTL confirmation. **(A)** The four types of populations used in the experiment: recombinant inbred lines (RILs), the ancestral strain CB4856, the ancestral strain N2, and introgression lines (ILs). These were exposed to three separate treatments: control (48 h at 20°C), heat-stress (46 h at 20°C followed by 2 h at 35°C), and recovery (as heat-stress, followed by an additional 2 h at 20°C). After the experiment, populations of the strains were harvested and used for transcriptome measurement by microarray. **(B)** the percentage of eQTL confirmed based on differentially expressed genes per IL and per RIL (FDR = 0.05). The percentages are split-out over the three environments, and the two eQTL type (*cis* or *trans*). The eQTL were mapped using the RIL population, hence that population is expected to confirm the eQTL to a high degree. **(C)** An example of one *cis*-eQTL (for gene F26E4.1) and one *trans*-eQTL (for gene K06H6.1) in the IL population. On the *x*-axis the genotype at the QTL locus is plotted, and on the *y*-axis the expression (R_{N2}).

assumed that the measurement error (σ_e^2) was similar for both populations and calculated the heritability ratio between both populations (**Figure 3A**). Based on the additive simulation, we expected 95% of the variance ratios to fall between −3.6 and −1.7, whereas under a model with closely-linked QTL that cancel each other out in the same genetic background (balanced), the 95% interval falls between −24.9 and 4.9. The balanced closely-linked QTL architecture was modeled such that QTL within such a cluster would only separate occasionally by a recombination in the RIL and IL population. The shape of the distribution of genes with significant heritability seems to form a combination of additive and balanced closely linked QTL (**Figure 3A**). To get an estimate of the contribution of the two models, we used the difference in the additive and balanced closely linked distributions from the simulation to estimate how the traits are divided over the two QTL models (**Figure 3B**). Analysis of the two distributions revealed that in all three environments additivity was the most likely QTL model (60–93% of genes with significant H^2). Balanced closely linked QTL explained up to 40% of relative heritability in the control environment, but only 7% in the heat stress and recovery environments (**Supplementary Table 5**). Therefore, we hypothesize that the balanced closely linked QTL model is an important eQTL architecture in the nematode *C. elegans*.

Two Predictions From the Balanced Closely Linked eQTL Microarchitecture

The balanced closely linked QTL model comes with two testable predictions: (i) the genes with heritable variation in gene expression were not differentially expressed in the parental

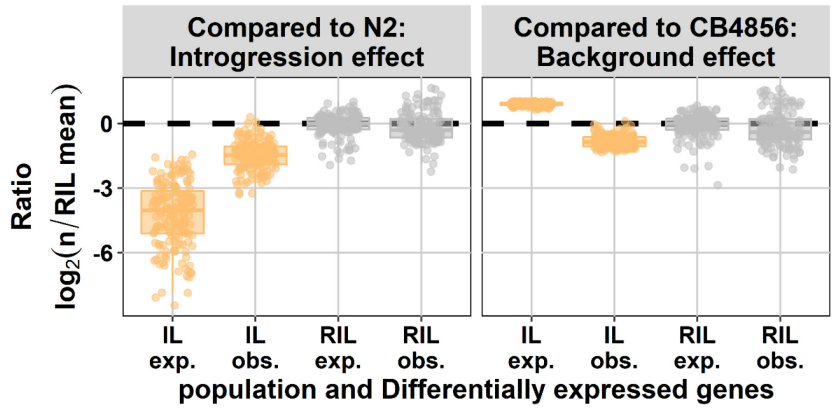

FIGURE 2 | Number of differentially expressed genes (DEG) expected (exp.) from eQTL mapping and observed experimentally (obs.). On the *x*-axis, the population type is shown and on the *y*-axis, the expected and observed ratio of differentially expressed genes is shown, both normalized for the expectation in the RILs [eQTL threshold, $-\log_{10}(p) > 3.9$; DEG threshold, $-\log_{10}(p) > 5$]. In the left panel the effect compared to N2 is shown (effect of the CB4856 loci; introgressions in the IL panel) and in the right panel the effect compared to CB4856 is shown (effect of the N2 loci; genetic background in the IL panel).

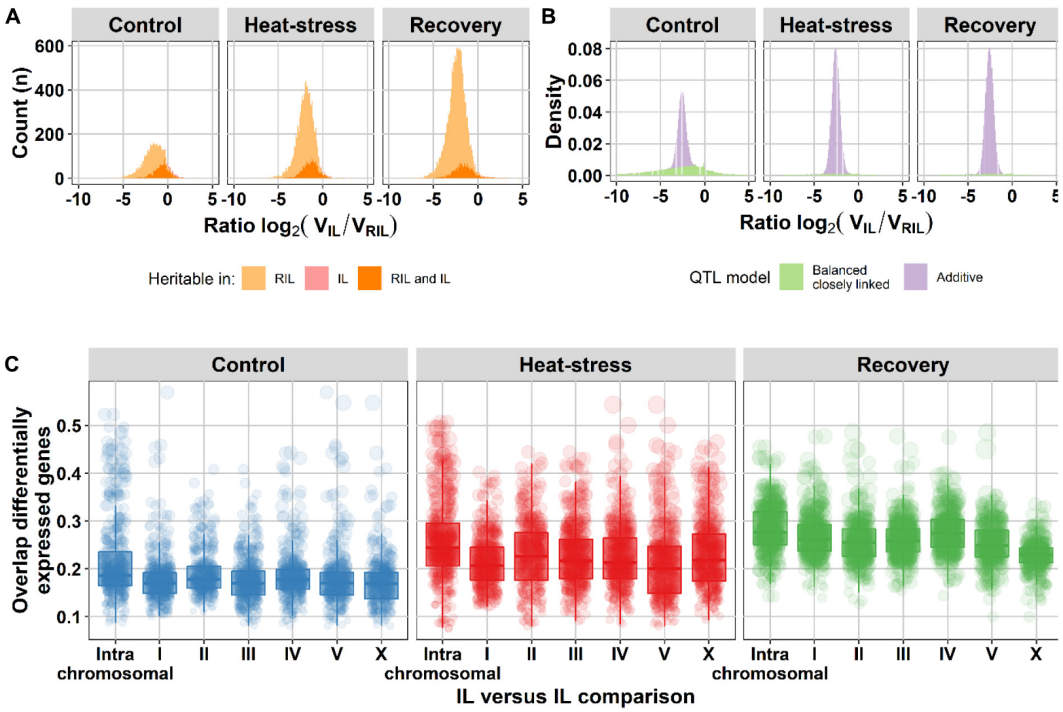

FIGURE 3 | Balanced closely linked QTL can explain heritability ratios between the RIL and IL populations. **(A)** Histogram of the heritability ratio in the IL versus the RIL population of genes with significant gene expression heritability (permutation, FDR = 0.05). The light orange surface indicates microarray spots of genes with heritable variation in gene expression only in the RIL population (3,262 in control, 7,545 in heat-stress, and 12,176 in recovery), the pink surface microarray spots of genes with only heritable variation in gene expression in the IL population (145, 59, and 71, respectively) and the orange color microarray spots of genes with heritable variation in gene expression in both populations (1,216, 1,588, and 1,507 respectively). **(B)** The density distribution of the heritability ratios of genes with heritable gene expression variation over the balanced closely linked (green) and the additive (purple) QTL model. Trait models consist only of additive or balanced closely linked QTL. **(C)** Overlap in differentially expressed genes between ILs [threshold $-\log_{10}(p) > 5$]. The comparisons are grouped per chromosome (*x*-axis): the intra-chromosomal comparisons and all other chromosomes versus chromosome I, II, III, IV, V, or X. On the *y*-axis the fraction overlap is shown. The dots represent one IL versus IL comparison the size corresponds to the number of overlapping differentially expressed genes.

strains, (ii) differential expression of genes was not only linked to one location in ILs but to clusters of ILs perturbing distant loci (multiple balanced clusters).

First, we calculated the number of genes with significant heritabilities. For the RILs, 4,478 spots (2,910 genes) displayed significant heritability (FDR = 0.05) in control environment

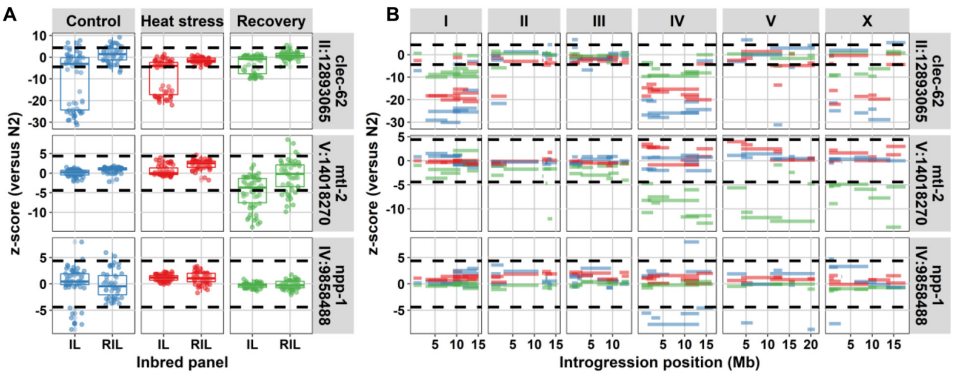

FIGURE 4 | Three genes with high heritability in gene expression variation in the IL population compared to the RIL population. **(A)** Comparison in expression between IL and RIL population for *clec-62*, *mtl-2*, and *npp-1*. On the *x*-axis, the population type and on the *y*-axis, the *z*-score with the N2 parental lines. The dashed horizontal lines indicate the significance threshold [*z*-score, $-\log_{10}(p) > 5$]. **(B)** The expression per introgression. On the *x*-axis, the introgression location is shown per chromosome. On the *y*-axis, the *z*-score with the N2 parental line is shown. The dashed horizontal lines indicate the significance threshold [*z*-score, $-\log_{10}(p) > 5$]. The colors indicate the environments (blue for control, red for heat-stress, and green for recovery).

($H^2 \geq 0.68$), 9,133 spots (5,072 genes) in heat-shock environment ($H^2 \geq 0.70$), and 13,683 spots (7,537 genes) in recovery environment ($H^2 \geq 0.62$). For the ILs, 1,361 spots (1,012 genes) displayed significant heritability (FDR = 0.05) in control environment ($H^2 \geq 0.70$), 1,647 spots (1,123 genes) in heat-shock environment ($H^2 \geq 0.71$), and 1,578 spots (1,092 genes) in recovery environment ($H^2 \geq 0.64$). Then, we compared these with differential expression in the parental strains. Indeed, only 9.1–15.4% of the genes with heritable variation in gene expression were differentially expressed when comparing the two parental strains (FDR = 0.1; **Supplementary Table 6**). Also, the second prediction was found to be correct, on average 21.9% of the differentially expressed genes were shared between IL strains covering loci on different chromosomes, which was only slightly less than the average of 25.1% between ILs covering the same chromosome (**Figure 3C** and **Supplementary Figure 7**, and for examples **Figure 4**). Therefore, we concluded that localized genetic complexity was a major part of the eQTL architecture.

Furthermore, differences in heritability between the RIL and IL populations over environments indicated that the control environment was different from heat-shock and recovery. The balanced closely linked architecture, which predicted a broad range of relative heritability, seemed to be most prominent in the control environment. It could be that these complex architectures were more important for the developmental process. This environmental effect could be illustrated by the expression variation of the genes: *mtl-2*, *npp-1*, and *clec-62*. Of which *mtl-2* and *npp-1* showed an interaction with the environment. The gene *npp-1* only showed this interaction in the control environments, environment where ILs on multiple chromosomes showed differential expression of this gene, yet for *mtl-2* ILs with differential expression on different chromosomes were found during recovery conditions, whereas *clec-62* shows ILs with differential expression on different chromosomes in control and heat stress conditions (**Figure 4**). The nuclear core complex protein NPP-1 is involved in spindle formation and important for oogenesis, a process that was interrupted during heat-shock (Schetter et al., 2006; Jovic et al., 2017). MTL-2, one of two metallothioneins in *C. elegans*, was expressed in the intestine and was induced after heat-shock (Freedman et al., 1993). The pattern observed for *clec-62* was exceptional as it showed a consistent response over all three environments where IL strains with an introgression on chromosome I, chromosome IV, and chromosome X showed much lower expression compared to the N2 ancestor. Furthermore, also an IL strain with an introgression on chromosome II and one with an introgression on chromosome V displayed this phenotype. C-type lectin 62 is a gene belonging to the extensive *clec*-family, which is thought to be involved in *C. elegans* immunity (Schulenburg et al., 2008) although its exact function is not known.

DISCUSSION

Most eQTL Mapped in RILs Are Replicable in ILs

Here, we show that eQTL mapped in a RIL population can be confirmed in an IL population. We found that the majority of *cis*-eQTL are replicable in the IL population (on average 69.5% per IL), whereas *trans*-eQTL are less replicable in ILs (on average 42.2% per IL). This likely reflects the mainly monogenic architecture of *cis*-eQTL, versus the more polygenic and environment-dependent *trans*-eQTL (Li et al., 2006; Keurentjes et al., 2007; Rockman et al., 2010; Snoek et al., 2017; Albert et al., 2018). Especially the confirmation of *trans*-bands by ILs shows that these regulatory hot-spots are robust regardless of genetic background. It should be noted that this study does have limited power per introgression line as each line was only measured once per condition. Hence we avoid conclusions based on a single ILs throughout this paper. However, the tiling nature of the population does mean that each locus is covered by multiple ILs (Doroszuk et al., 2009). Furthermore, the RIL and IL experiment were separated in time, which could lead to unexpected effects influencing the replication.

This study helps the search for candidate genes underlying these *trans*-bands, by narrowing-down the regions. However, additional experiments and replicates are recommendable and currently being pursued for the heat-stress *trans*-band on chromosome IV. Furthermore, it shows how ILs can be used to narrow-down these eQTL hotspots (Snoek et al., 2012), using correlation analysis previously used to link *trans*-bands to genes and biological processes (Andersen et al., 2014; Sterken et al., 2017). These findings show on a large scale that QTL mapped using a single marker model in a moderately sized RIL population are reliably replicable in a population with a different genetic structure, which confirms findings for single traits reported in *C. elegans* and beyond [for example, see (Snoek et al., 2012; Andersen et al., 2014; Gao et al., 2018)].

Introgression Lines Indicate the Presence of Parental-Balanced, Polygenic, Traits

The difference in genetic complexity between RILs and ILs can be leveraged to understand trait architectures. Despite confirming many QTL, it has been noted that ILs often tell a different story than RILs, implying abundant genetic interactions, not uncovered by RILs [as reviewed by Mackay (2014)]. In this study, we also show that the number of differentially expressed genes in the N2 genetic background of the ILs is lower than expected compared to RIL-based estimations. Furthermore, eQTL mapped in the RIL population also led to an underestimation of the number of differentially expressed genes due to the introgressions in the ILs. To summarize, the introgressions display more-than-additive effects, whereas the genetic background shows less-than-additive effects.

Our findings are in line with observations in introgression lines in *C. elegans*, showing that for some traits more and different QTL than expected from RILs can be found (e.g., Gaertner et al., 2012; Glater et al., 2014; Snoek et al., 2014a). Additionally, our findings are also in line with findings in other organisms, where such effects have been reported for different types of traits (Shao et al., 2008; Gale et al., 2009; Spiezio et al., 2012). In this study, we show this effect in general, over many gene expression traits. Moreover, a direct comparison of trait mapping in a genome-wide IL population versus a RIL population has only been conducted in a few studies, as far as we are aware (Glater et al., 2014; Snoek et al., 2014a). It has been argued that ILs showing more- or less-than-additive effects compared to the parental strains is a hallmark of epistasis [as reviewed by Mackay (2014)]. However, recent studies in yeast show that additivity underlies most of the heritable trait variation among inbred lines, where epistasis accounts for phenotypic extremes (Bloom et al., 2015; Forsberg et al., 2017; Albert et al., 2018). Thus remains the question, what kind of trait architecture drives our observations: pervasive epistasis or widespread additivity?

The Developmental Trait Architecture Consists of Balanced Closely Linked QTL

Trait architectures underlying natural variation in *C. elegans* differ strongly over traits. For example, currently 25 quantitative trait nucleotides (QTNs) are known in *C. elegans*, capturing a majority of the heritable variation for particular traits [reviewed by Gaertner and Phillips (2010) and Rockman (2012) and studies by Andersen et al. (2014); Noble et al. (2015), Schmid et al. (2015); Cook et al. (2016), Greene et al. (2016a; 2016b), Large et al. (2016); Ben-David et al. (2017), Sterken et al. (2017); Zdraljevic et al. (2017, 2019), Hahnel et al. (2018)]. However, there are also examples of traits that are highly heritable but have only yielded complex or few QTL (let alone QTNs). For example, a study on bacterial preference of *C. elegans* noted a relatively high heritability (0.46) for *Serratia marcescens* over *E. coli*, yet uncovered only a single QTL in the RILs used for mapping whereas multiple were expected (Glater et al., 2014). Furthermore, recent work from our group, studying metabolite abundances showed high heritability (from 0.32 up to 0.82) corresponding to a few uncovered QTL (Gao et al., 2018). Intriguingly, several studies imply trait architectures consisting of closely linked QTL that are balanced in the parental strains (Gaertner et al., 2012; Glater et al., 2014; Bernstein et al., 2019). By simulation, we obtained evidence that some trait architectures can be differentiated by relative heritability between IL and RIL populations which is a property we can reliably measure in our two populations, as it only relies on determining trait variance. In this way, we found that for genes without an eQTL, but showing high heritability, the most parsimonious explanation lies in an architecture comprised of (multiple) closely linked QTL clusters. This type of architecture seems especially prominent during normal development where it could affect approximately 40% of genes with heritable expression variation. It should be noted that only strictly balanced closely linked and strictly additive QTL distributions were modeled. Furthermore, a study containing more replication to take in account intra-strain variance should be conducted to confirm this hypothesis. This would also benefit from a better knowledge on the trait architecture based on RILs, as in Bernstein et al. (2019). We think the observed distribution suggests that a combination of (both or multiple) types of QTL make up the complete genetic architecture of a trait.

In recent years, the number of parameters known to affect gene expression variation in the context of natural variation has steadily grown. It is currently clear that environment (Li et al., 2006; Snoek et al., 2017), age (Vinuela et al., 2010), development (Francesconi and Lehner, 2014), and background mutations (Sterken et al., 2017) affect gene expression variation and eQTL distribution in *C. elegans*. These factors could also contribute to the observed heritability differences between ILs and RILs. For environment, we can confirm that it plays an important role: the heritability ratio is strongly dependent on environment. This is in line with effects observed for *trans*-eQTL, which are also strongly environment dependent, implying heritability is as well (Li et al., 2006; Snoek et al., 2017). Developmental differences between the strains in each population could also drive some of the observed heritability ratios. For example, ILs could be developmentally more homogenous than RILs, which could result in lower levels of trait variance in ILs. It is clear that age affects *trans*-eQTL and heritability (Vinuela et al., 2010, 2012). However, it is unclear how that would affect the heritability ratios we measure in general; it seems unlikely that developmental differences result in a

specific variance signature as changes occur in both populations. More specific, in this experiment strains are tightly synchronized and developmental stage is observed to be L4 before isolation, therefore we expect that these potential sources of increased variance are controlled. Hence, the relative heritabilities under normal development imply clusters of closely linked QTL.

The Role of Cryptic Genetic Variation on Trait Complexity

We used the contrast in genetic complexity of the RIL and IL population to investigate trait complexity. The relative heritabilities between these populations differed over the three environments. In our previous study on the RILs, we show that the *trans*-eQTL architecture is different over the three environments, showing that *trans*-eQTL contribute most to cryptic genetic variation. Here, we find that the micro-architectures is more complex in an ambient environment.

Development of an organism is a tightly regulated process and completion of development into a reproducing adult is essential for reproduction, hence fitness. The nematode *C. elegans* experiences outbreeding depression (Dolgin et al., 2007), given the strains N2 and CB4856 are genetically distinct, it is likely that inbred lines constructed from these two strains disrupt some linked loci involved in this process (Seidel et al., 2008; Snoek et al., 2014a). We hypothesize that especially the ambient environment stimulating normal development is prone to accentuate the effects of genetic 'mismatches' resulting in trait variation. First, local sites are expected to co-segregate over many generations in *C. elegans*, making it likely that they are filled with compensatory mutations (Rockman et al., 2010). Second, development is a polygenic, tightly regulated, process. On the level of gene expression, there are strong changes over the entire development from egg to adult, and during the last juvenile stage (L4) especially (Francesconi and Lehner, 2014; Hendriks et al., 2014; Snoek et al., 2014b; Jovic et al., 2019b). Together, these two effects may cause the *C. elegans* populations in ambient environments to show micro-architectures that are more complex.

On the other hand, stress responses are also tightly regulated, but initially aimed at reaching a stress-survival mode, which is a process that involves fewer genes than development. The heat-shock response in *C. elegans* starts by de-phosphorylation of HSF-1 and DAF-16 and entry of these transcription factors into the nucleus [as reviewed by Rodriguez et al. (2013)]. Once there, they regulate the expression of chaperones that function to minimize cellular damage. The physiological effects of heat-shock exposure are duration dependent, where a 2-h exposure in our setup does not result in severe phenotypic effects (Jovic et al., 2017). However, longer exposures lead to less movement, delayed egg-laying, and early death. Between N2 and CB4856, there are x temperature related traits mapped: thermal preference (mapping to the left of chromosome X) (Gaertner et al., 2012), recovery from heat-stress (mapping to the center-right of chromosome II) (Rodriguez et al., 2012), reproduction after heat-stress (mapping to the center-right of chromosome II and the left of chromosome IV) (Rodriguez et al., 2012), and temperature-related size differences (mapping to *tra-3* on chromosome IV) (Kammenga et al., 2007). The chromosome II region associated with recovery from heat-stress was confirmed by ILs WN225 and WN226 in the original paper, and falls under a heat-stress *trans*-band (Snoek et al., 2017). Here, we find that both ILs confirm the existence of the *trans*-band. Hence, the gene expression affected by a polymorphic gene on chromosome II could be linked to recovery from heat-stress.

Overall, our results show that populations in both the heat-shock and recovery from heat-shock environments show fewer QTL with complex micro-architectures. We hypothesize that this is due to the effect of the strong transcriptional response induced by heat shock (Brunquell et al., 2016; Jovic et al., 2017). It is possible that this abrupt disturbance emphasizes regulatory variation in only a few key response pathways compared to the more subtle and complex process of development.

Implications for Understanding Life-History Adaptations

Our data facilitate exploring the functional implications of gene transcriptional changes following heat-shock and recovery. The gene transcriptional landscape consists of multiple gene networks underlying different traits, including life-history traits (Valba et al., 2015). Gene transcriptional response to stress and subsequent recovery are likely to shape life-histories given that many of the involved polymorphic genes may have clear functionalities. For instance, polymorphic regulators in genes affecting thermo-sensation may be important for controlling locomotor behavior induced by thermal stress via control neural decision-making (Stegeman et al., 2019). *C. elegans* is ectothermal and inhabits ecological niches that are prone to rapid temperature fluctuations in Europa (N2), but stable temperatures in Hawaii (CB4856) (Crombie et al., 2019). Hence, in Europe it needs to cope with these sudden temperature changes. Jiang et al. (2018) studied the genetics of cold-warm changes in *C. elegans* and detected ISY-1 and ZIP-10 as major gate keepers of temperature change responses (Jiang et al., 2018). It would be interesting to investigate if polymorphic gene expression regulators in this warm-cold switch are likely to underlie these temperature switches. Moreover, more variable alleles affecting gene expression could be detected using the recently developed multi parental RIL populations as they show variation in life history traits (Noble et al., 2017; Snoek et al., 2019). We recently showed that thermotolerance in *C. elegans* could be predicted from the gene expression resilience patterns in worms exposed to heat stress (Jovic et al., 2019a,b). It was found that the predictive outcomes also hold up across different genotypes, suggesting that polymorphic regulators play an important role in thermotolerance.

CONCLUSION

We present an eQTL experiment conducted with ILs in three environments covering the same conditions as a RIL experiment published previously (Snoek et al., 2017). We show that ILs can replicate both *cis*- and *trans*-eQTL; furthermore, we present evidence supporting 14 eQTL hot spots (*trans*-bands). Yet, eQTL

mapped in the RIL population systematically under-estimate the impact of a single introgression on the transcriptome and over-estimate the impact of the genetic background, suggesting additional genetic complexity. We present evidence that during normal growth multiple clusters of closely linked QTL could underly additional genetic complexity.

Further understanding of this phenomenon requires systematic dissection of many QTL – a process that in this study was only undertaken for *trans*-bands – using introgression lines. It remains to be determined what kind of role epistasis plays in trait variation. Currently, RIL populations in most species are typically of insufficient size to detect any epistasis beyond two loci interactions. Careful crosses with ILs to generate double-loci ILs might be a way forward to further our understanding of trait regulation in the context of natural genetic variation.

DATA AVAILABILITY STATEMENT

The transcriptome datasets used in the analysis are deposited at ArrayExpress E-MTAB-5779 and E-MTAB-7424.

AUTHOR CONTRIBUTIONS

MS, BS, and JK conceived and designed the experiments. MS wrote the manuscript, with input from JK and BS. MS, RB, RV, and JR performed the experiments. MS analyzed the data, with input from BS. All authors commented on the manuscript.

FUNDING

BS was funded by ERASysbio-plus ZonMW project GRAPPLE – Iterative modeling of gene regulatory interactions underlying stress, disease and aging in *C. elegans* (project 90.201.066) and Netherlands Organisation for Scientific Research (project no. 823.01.001). JK was funded by NIH grant 1R01AA 026658-01. The funding bodies had no role in the design nor the collection, analysis, and interpretation of the data, nor the writing of the manuscript.

ACKNOWLEDGMENTS

We thank the GRAPPLE project partners: Olga Valba, Sreenivas Chavali, Benjamin Lang, Mirko Francesconi, Rachel Brenchley, Arjen van't Hof, Sergei Nechaev, Olga Vasieva, Madan Babu, Andrew Cossins, and Ben Lehner. We thank Harm Nijveen for assistance with WormQTL2.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.501376/full#supplementary-material

Supplementary Figure 1 | Coverage per locus and per QTL. **(A)** The coverage in CB4856 loci per location on the genome, split out for ILs and RILs. The 56 ILs together have a higher coverage over the chromosome arms, where also most QTL map. The 48 RILs have a more homogenous distribution, only at the *peel-1/zeel-1* locus on chromosome I there is low coverage (Seidel et al., 2008). **(B)**. A histogram of the number of CB4856 loci covering an eQTL. Typically, an eQTL is covered by CB4856 loci of 2 ILs and 23 RILs (median).

Supplementary Figure 2 | Confirmation of eQTL by significance thresholds and correlation analysis. **(A)** The percentage of confirmed eQTL per population per eQTL-type. On the *x*-axis, the $-\log_{10}(p)$ significance threshold in the RILs is shown (binned per 0.5 in the analysis) and on the *y*-axis the percentage of confirmed eQTL per bin per strain is shown (at least five confirmable eQTL required per bin). For example, for the RIL WN98 under recovery conditions there were 115 *cis*-eQTL with a significance between 4.0 and 4.5 in its CB4856 regions, 91 of these were confirmed based on the *z*-score (79.1%). This example results in a dot where the size corresponds to the number of eQTL and the color indicates the environments. The line shown is a linear regression curve, where the surrounding gray area is the confidence interval. **(B)** The Pearson correlation of eQTL effects with gene expression in the ILs and the RILs. On the *x*-axis the population type is plotted and on the *y*-axis the Pearson correlation. The correlations are split-out over the three environments (control, heat stress, and recovery) and over eQTL type (*cis* or *trans*).

Supplementary Figure 3 | Differentially expressed genes in RILs and ILs. **(A)** Number of differentially expressed genes (DEG) expected from eQTL mapping and observed experimentally, split out per environment. On the *x*-axis the population and on the *y*-axis the expected and observed ratio of differentially expressed genes is shown, both normalized for the expectation in the RILs [eQTL threshold, $-\log_{10}(p) > 3.9$; DEG threshold, $-\log_{10}(p) > 5$]. In the top panel the effect compared to N2 is shown (effect of the CB4856 loci; introgressions in the IL panel) and in the bottom panel the effect compared to CB4856 is shown (effect of the N2 loci; genetic background in the IL panel). **(B)** The effect of different thresholds for eQTL significance (*x*-axis) and DEG calling (*y*-axis) on the ratio between observed and expected.

Supplementary Figure 4 | Genes with heritable variation in gene expression in the RIL and IL populations. **(A)** A histogram of the measured heritabilities for the IL and RIL populations, negative values have been discarded. Colors indicate whether the heritability was significant (FDR = 0.05) in the control (blue), heat-stress (red), or recovery (green) environment. **(B)** The overlap in genes with significant heritability in gene expression variation over the environments for the RIL population (FDR = 0.05). **(C)** The overlap in genes with significant heritable variation in gene expression over the environments for the IL population (FDR = 0.05).

Supplementary Figure 5 | A graphical overview of the modeled trait architectures. The arrows indicate QTL effect sizes as found in the N2 genotype (orange) and the CB4856 genotype (blue). Perfectly opposed arrows, one of which orange and the other blue, indicate that the same QTL has an opposite effect in the two genotypes (e.g., as seen for Additive random distribution). When an arrow lacks a perfectly opposed arrow of another color, it means that the QTL is only found in that particular genotype. Epistatic interactions are indicated as a line connecting two rectangles.

Supplementary Figure 6 | Variance ratios between IL and RIL population for the simulated trait architectures (**Supplementary Figure 5**). The density of occurrence of variance ratios per simulation is given. The color scale indicates how many QTL for a trait were simulated.

Supplementary Figure 7 | Overlap in differentially expressed genes between ILs [threshold $-\log_{10}(p) > 5$]. The ILs are ordered based on introgression on the *x*- and *y*-axis. The location of the CB4856 segment is indicated behind the IL strain-code. The fraction overlap shown is calculated as the percentage of unique differentially expressed genes in the two compared ILs. The overlap is shown for the control environment **(A)**, the heat-stress environment **(B)**, and the recovery environment **(C)**.

Supplementary Table 1 | The genotypes of the strains used in this study, −1 denotes a CB4856 genotype, a 1 denotes an N2 genotype. The genotypes are given per chromosome and per position (based on WormBase version WS258).

Supplementary Table 2 | Table of the eQTL with gene expression comparisons per strain per environment. The table lists the strain name, strain type and the comparisons are split out for the *cis*- and *trans*-eQTL. The correlation reported is the Pearson correlation and its significance. The number of eQTL present in the CB4856 loci is given, as is the percentage that is also differentially expressed (FDR = 0.05).

Supplementary Table 3 | Summary of the comparison of gene expression in the ILs with the *trans*-bands identified previously (Snoek et al., 2017). The location of the *trans*-band and the number of genes and microarray spots affected are listed. The genotypes of the ILs with an introgression adjacent to – or covering the *trans*-band are listed. The genotypes of the ILs that confirm the *trans*-band and the *trans*-band locus that these ILs imply are also given.

Supplementary Table 4 | Table with the heritabilities per spot per environment per population. The estimated genetic variance (V_g) and the technical variance (measured from the replicated parental lines –measurement error – V_e) are given, as is the heritability (H^2). The FDR0.05 column indicates the threshold of significance based on permutation.

Supplementary Table 5 | A table with the estimated fraction and amount of traits for an additive QTL architecture and a balanced closely-linked QTL architecture. The numbers are calculated based on only a single type of QTL occurring for a particular architecture. For example, additive means *n* QTL that display an additive effect, without any other type of effect for that simulated trait.

Supplementary Table 6 | Differentially expressed genes between N2 and CB4856 for the three environments: control, heat-stress, and recovery from heat stress. The columns in the table: SpotID, the Agilent spot identifier, the information on the gene expression detected by that spot is given [chromosome, location (start/stop), strand, and three identifiers]. Two columns show if genes are significantly heritable; general (in any population) and specific (which population, or both). The significance and effect columns give the output of the linear model when comparing CB4856 versus N2 (positive effect indicates the gene is higher expressed in N2). The last column gives the significance as corrected for multiple testing (FDR).

Supplementary File 1 | Comparison of gene expression in ILs with eQTL in *trans*-bands identified in the RIL population. Per *trans*-band, a figure was constructed. In **(A)** the name of the *trans*-band is given at the top of the panel (e.g., chromosome I, 3.5–4 million bases) the correlation of the relative gene expression (R_{N2}) was correlated with the eQTL effect for the ILs covering or nearby the *trans*-band. Significance was determined based on the highest correlation in the N2 strains, only strains with a stronger correlation than N2 were scored as significant. In **(B)** the genetic map of the ILs of the *trans*-band chromosome is shown. On the *x*-axis the genomic location (in million bases) is shown and on the *y*-axis the ILs. In **(C)** the actual correlation between expression in the ILs and eQTL effect in the RILs is shown. With on the *x*-axis the eQTL effect and on the *y*-axis R_{N2}. Each dot represents a spot on the microarray with an eQTL in the *trans*-band investigated. The line indicates the estimated slope and the pale blue area around the line indicates the confidence interval of the fit.

REFERENCES

Albert, F. W., Bloom, J. S., Siegel, J., Day, L., and Kruglyak, L. (2018). Genetics of trans-regulatory variation in gene expression. *eLife* 7:e35471.

Albert, F. W., and Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.* 16, 197–212. doi: 10.1038/nrg3891

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. doi: 10.1371/journal.pgen.1004156

Ben-David, E., Burga, A., and Kruglyak, L. (2017). A maternal-effect selfish genetic element in *Caenorhabditis elegans*. *Science* 356, 1051–1055. doi: 10.1126/science.aan0621

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Bernstein, M. R., Zdraljevic, S., Andersen, E. C., and Rockman, M. V. (2019). Tightly linked antagonistic-effect loci underlie polygenic phenotypic variation in *C. elegans*. *Evol. Lett.* 3, 462–473. doi: 10.1002/evl3.139

Bloom, J. S., Kotenko, I., Sadhu, M. J., Treusch, S., Albert, F. W., Kruglyak, L., et al. (2015). Genetic interactions contribute less than additive effects to quantitative trait variation in yeast. *Nat. Commun.* 6:8712.

Brem, R. B., and Kruglyak, L. (2005). The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 102, 1572–1577. doi: 10.1073/pnas.0408709102

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

Brunquell, J., Morris, S., Lu, Y., Cheng, F., and Westerheide, S. D. (2016). The genome-wide role of HSF-1 in the regulation of gene expression in *Caenorhabditis elegans*. *BMC Genomics* 17:559. doi: 10.1186/s12864-016-2837-5

Conover, W. J., Johnson, M. E., and Johnson, M. M. (1981). A comparative-study of tests for homogeneity of variances, with applications to the outer continental-shelf bidding data. *Technometrics* 23, 351–361. doi: 10.1080/00401706.1981.10487680

Cook, D. E., Zdraljevic, S., Tanny, R. E., Seo, B., and Riccardi, D. D. (2016). The genetic basis of natural variation in *Caenorhabditis elegans* telomere length. *Genetics* 204, 371–383.

Crombie, T. A., Zdraljevic, S., Cook, D. E., Tanny, R. E., and Brady, S. C. (2019). Deep sampling of Hawaiian *Caenorhabditis elegans* reveals high genetic diversity and admixture with global populations. *eLife* 8: e50465.

Dolgin, E. S., Charlesworth, B., Baird, S. E., and Cutter, A. D. (2007). Inbreeding and outbreeding depression in Caenorhabditis nematodes. *Evolution* 61, 1339–1352. doi: 10.1111/j.1558-5646.2007.00118.x

Doroszuk, A., Snoek, L. B., Fradin, E., Riksen, J., and Kammenga, J. (2009). A genome-wide library of CB4856/N2 introgression lines of *Caenorhabditis elegans*. *Nucleic Acids Res.* 37:e110. doi: 10.1093/nar/gkp528

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. doi: 10.1371/journal.pbio.1001230

Forsberg, S. K., Bloom, J. S., Sadhu, M. J., Kruglyak, L., and Carlborg, O. (2017). Accounting for genetic interactions improves modeling of individual quantitative trait phenotypes in yeast. *Nat. Genet.* 49, 497–503. doi: 10.1038/ng.3800

Fournier, T., and Schacherer, J. (2017). Genetic backgrounds and hidden trait complexity in natural populations. *Curr. Opin. Genet. Dev.* 47, 48–53. doi: 10.1016/j.gde.2017.08.009

Francesconi, M., and Lehner, B. (2014). The effects of genetic variation on gene expression dynamics during development. *Nature* 505, 208–211. doi: 10.1038/nature12772

Freedman, J. H., Slice, L. W., Dixon, D., Fire, A., and Rubin, C. S. (1993). The novel metallothionein genes of *Caenorhabditis elegans*. Structural organization and inducible, cell-specific expression. *J. Biol. Chem.* 268, 2554–2564.

Gaertner, B. E., Parmenter, M. D., Rockman, M. V., Kruglyak, L., and Phillips, P. C. (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. doi: 10.1534/genetics.112.142877

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. doi: 10.1017/s0016672310000601

Gale, G. D., Yazdi, R. D., Khan, A. H., Lusis, A. J., and Davis, R. C. (2009). A genome-wide panel of congenic mice reveals widespread epistasis of behavior quantitative trait loci. *Mol. Psychiatry* 14, 631–645. doi: 10.1038/mp.2008.4

Gao, A. W. W., Sterken, M. G., de Bos, J. U., van Creij, J., and Kamble, R. (2018). Natural genetic variation in *C. elegans* identified genomic loci controlling metabolite levels. *Genome Res.* 28, 1296–1308. doi: 10.1101/gr.232322.117

Gilad, Y., Rifkin, S. A., and Pritchard, J. K. (2008). Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet.* 24, 408–415. doi: 10.1016/j.tig.2008.06.001

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* 4, 265–276. doi: 10.1534/g3.113.008649

Greene, J. S., Brown, M., Dobosiewicz, M., Ishida, I. G., and Macosko, E. Z. (2016a). Balancing selection shapes density-dependent foraging behaviour. *Nature* 539, 254–258. doi: 10.1038/nature19848

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

Hahnel, S. R., Zdraljevic, S., Rodriguez, B. C., Zhao, Y., McGrath, P. T., Andersen, E. C., et al. (2018). Extreme allelic heterogeneity at a *Caenorhabditis elegans* beta-tubulin locus explains natural resistance to benzimidazoles. *PLoS Pathog.* 14:e1007226. doi: 10.1371/journal.ppat.1007226

Hendriks, G. J., Gaidatzis, D., Aeschimann, F., and Grosshans, H. (2014). Extensive oscillatory gene expression during *C. elegans* larval development. *Mol. Cell* 53, 380–392. doi: 10.1016/j.molcel.2013.12.013

Jansen, R. C., and Nap, J. P. (2001). Genetical genomics: the added value from segregation. *Trends Genet.* 17, 388–391. doi: 10.1016/s0168-9525(01)02310-1

Jiang, W., Wei, Y., Long, Y., Owen, A., and Wang, B. (2018). A genetic program mediates cold-warming response and promotes stress-induced phenoptosis in *C. elegans*. *eLife* 7:e35037.

Jimenez-Gomez, J. M., Corwin, J. A., Joseph, B., Maloof, J. N., and Kliebenstein, D. J. (2011). Genomic Analysis of QTLs and genes altering natural variation in stochastic noise. *PLoS Genet.* 7:e1002295. doi: 10.1371/journal.pgen.1002295

Jimenez-Gomez, J. M., Wallace, A. D., and Maloof, J. N. (2010). Network analysis identifies ELF3 as a QTL for the shade avoidance response in *Arabidopsis*. *PLoS Genet.* 6:e1001100. doi: 10.1371/journal.pgen.1001100

Jovic, K., Grilli, J., Sterken, M. G., Snoek, B. L., and Riksen, J. A. G. (2019a). Transcriptome dynamics predict thermotolerance in *Caenorhabditis elegans*. *bioRxiv [Preprint]* doi: 10.1101/661652

Jovic, K., Grilli, J., Sterken, M. G., Snoek, B. L., and Riksen, J. A. G. (2019b). Transcriptome resilience predicts thermotolerance in *Caenorhabditis elegans*. *BMC Biol.* 17:102. doi: 10.1186/s12915-019-0725-6

Jovic, K., Sterken, M. G., Grilli, J., Bevers, R. P. J., and Rodriguez, M. (2017). Temporal dynamics of gene expression in heat-stressed *Caenorhabditis elegans*. *PLoS One* 12:e0189445. doi: 10.1371/journal.pone.0189445

Kammenga, J. E., Doroszuk, A., Riksen, J. A. G., Hazendonk, E., and Spiridon, L. (2007). A *Caenorhabditis elegans* wild type defies the temperature-size rule owing to a single nucleotide polymorphism in tra-3. *PLoS Genet.* 3:e30034. doi: 10.1371/journal.pgen.0030034

Keurentjes, J. J. B., Fu, J. Y., Terpstra, I. R., Garcia, J. M., and van den Ackerveken, G. (2007). Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1708–1713. doi: 10.1073/pnas.0610429104

Large, E. E., Xu, W., Zhao, Y., Brady, S. C., and Long, L. (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. doi: 10.1371/journal.pgen.1006219

Li, Y., Lvarez, O. A. A., Gutteling, E. W., Tijsterman, M., and Fu, J. J. (2006). Mapping determinants of gene expression plasticity by genetical genomics in *C-elegans*. *PLoS Genet.* 2:e20222. doi: 10.1371/journal.pgen.0020222

Mackay, T. F. C. (2014). Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nat. Rev. Genet.* 15, 22–33. doi: 10.1038/nrg3627

Muenke, M., Schell, U., Hehr, A., Robin, N. H., Losken, H. W., Schinzel, A., et al. (1994). A common mutation in the fibroblast growth factor receptor 1 gene in Pfeiffer syndrome. *Nat. Genet.* 8, 269–274. doi: 10.1038/ng1194-269

Nica, A. C., and Dermitzakis, E. T. (2013). Expression quantitative trait loci: present and future. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368:20120362. doi: 10.1098/rstb.2012.0362

Noble, L. M., Chang, A. S., McNelis, D., Kramer, M., and Yen, M. (2015). Natural Variation in plep-1 causes male-male copulatory behavior in *C. elegans*. *Curr. Biol.* 25, 2730–2737. doi: 10.1016/j.cub.2015.09.019

Noble, L. M., Chelo, I., Guzella, T., Afonso, B., Riccardi, D. D., Ammerman, P., et al. (2017). Polygenicity and epistasis underlie fitness-proximal traits in the *Caenorhabditis elegans* multiparental experimental evolution (CeMEE) panel. *Genetics* 207, 1663–1685. doi: 10.1534/genetics.117.300406

Radwan, J., and Babik, W. (2012). The genomics of adaptation Introduction. *Proc. R. Soc. B Biol. Sci.* 279, 5024–5028.

Ranjan, A., Budke, J. M., Rowland, S. D., Chitwood, D. H., and Kumar, R. (2016). eQTL regulating transcript levels associated with diverse biological processes in tomato. *Plant Physiol.* 172, 328–340. doi: 10.1104/pp.16.00289

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., and Law, C. W. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007

Rockman, M. V. (2012). The Qtn program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution* 66, 1–17. doi: 10.1111/j.1558-5646.2011.01486.x

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. doi: 10.1126/science.1194208

Rodriguez, M., Snoek, L. B., De Bono, M., and Kammenga, J. E. (2013). Worms under stress: *C. elegans* stress response and its relevance to complex human disease and aging. *Trends Genet.* 29, 367–374. doi: 10.1016/j.tig.2013.01.010

Rodriguez, M., Snoek, L. B., Riksen, J. A. G., Bevers, R. P., and Kammenga, J. E. (2012). Genetic variation for stress-response hormesis in *C. elegans* lifespan. *Exp. Gerontol.* 47, 581–587. doi: 10.1016/j.exger.2012.05.005

Schadt, E. E., Monks, S. A., Drake, T. A., Lusis, A. J., and Che, N. (2003). Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422, 297–302. doi: 10.1038/nature01434

Schetter, A., Askjaer, P., Piano, F., Mattaj, I., and Kemphues, K. (2006). Nucleoporins NPP-1, NPP-3, NPP-4, NPP-11 and NPP-13 are required for proper spindle orientation in *C. elegans*. *Dev. Biol.* 289, 360–371. doi: 10.1016/j.ydbio.2005.10.038

Schmid, T., Snoek, L. B., Frohli, E., van der Bent, M. L., and Kammenga, J. (2015). Systemic regulation of RAS/MAPK Signaling by the Serotonin Metabolite 5-HIAA. *PLoS Genet.* 11:e1005236. doi: 10.1371/journal.pgen.1005236

Schulenburg, H., Hoeppner, M. P., Weiner, J. III, and Bornberg-Bauer, E. (2008). Specificity of the innate immune system and diversity of C-type lectin domain (CTLD) proteins in the nematode *Caenorhabditis elegans*. *Immunobiology* 213, 237–250. doi: 10.1016/j.imbio.2007.12.004

Seidel, H. S., Rockman, M. V., and Kruglyak, L. (2008). Widespread genetic incompatibility in *C. elegans* maintained by balancing selection. *Science* 319, 589–594. doi: 10.1126/science.1151107

Shao, H., Burrage, L. C., Sinasac, D. S., Hill, A. E., and Ernest, S. R. (2008). Genetic architecture of complex traits: large phenotypic effects and pervasive epistasis. *Proc. Natl. Acad. Sci. U.S.A.* 105, 19910–19914. doi: 10.1073/pnas.0810388105

Smyth, G. K., and Speed, T. (2003). Normalization of cDNA microarray data. *Methods* 31, 265–273. doi: 10.1016/s1046-2023(03)00155-5

Snoek, B. L., Sterken, M. G., Bevers, R. P. J., Volkers, R. J. M., and Van't Hof, A. (2017). Contribution of trans regulatory eQTL to cryptic genetic variation in *C. elegans*. *BMC Genomics* 18:500. doi: 10.1186/s12864-017-3899-8

Snoek, B. L., Sterken, M. G., Hartanto, M., van Zuilichem, A. J., Kammenga, J. E., de Ridder, D., et al. (2020). WormQTL2: an interactive platform for systems genetics in *Caenorhabditis elegans*. *Database* 2020:baz149.

Snoek, L. B., Orbidans, H. E., Stastna, J. J., Aartse, A., and Rodriguez, M. (2014a). Widespread genomic incompatibilities in *Caenorhabditis elegans*. *G3* 4, 1813–1823. doi: 10.1534/g3.114.013151

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

Snoek, L. B., Terpstra, I. R., Dekter, R., Van den Ackerveken, G., and Peeters, A. J. (2012). Genetical genomics reveals large scale genotype-by-environment interactions in *Arabidopsis thaliana*. *Front. Genet.* 3:317. doi: 10.3389/fgene.2012.00317

Snoek, B. L., Volkers, R. J. M., Nijveen, H., Petersen, C., Dirksen, P., Sterken, M. G., 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, 1–17. doi: 10.1186/s12915-019-0642-8

Spiezio, S. H., Takada, T., Shiroishi, T., and Nadeau, J. H. (2012). Genetic divergence and the genetic architecture of complex traits in chromosome substitution strains of mice. *BMC Genet.* 13:38. doi: 10.1186/1471-2156-13-38

Stegeman, G. W., Medina, D., Cutter, A. D., and Ryu, W. S. (2019). Neuro-genetic plasticity of *Caenorhabditis elegans* behavioral thermal tolerance. *BMC Neurosci.* 20:26. doi: 10.1186/s12868-019-0510-z

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. doi: 10.1016/j.tig.2015.02.009

Sterken, M. G., van Bemmelen, van der Plaat, L., Riksen, J. A. G., Rodriguez, M., and Schmid, T. (2017). Ras/MAPK Modifier Loci Revealed by eQTL in *Caenorhabditis elegans*. *G3* 7, 3185–3193. doi: 10.1534/g3.117.1120

R Core Team, (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team.

Terpstra, I. R., Snoek, L. B., Keurentjes, J. J., Peeters, A. J., and van den Ackerveken, G. (2010). Regulatory network identification by genetical genomics: signaling downstream of the Arabidopsis receptor-like kinase ERECTA. *Plant Physiol.* 154, 1067–1078. doi: 10.1104/pp.110.159996

Thompson, O. A., Snoek, L. B., Nijveen, H., Sterken, M. G., and Volkers, R. J. (2015). Remarkably Divergent Regions Punctuate the Genome Assembly of the *Caenorhabditis elegans* Hawaiian Strain CB4856. *Genetics* 200, 975–989. doi: 10.1534/genetics.115.175950

Valba, O. V., Nechaev, S. K., Sterken, M. G., Snoek, L. B., and Kammenga, J. E. (2015). On predicting regulatory genes by analysis of functional networks in *C. elegans*. *BioData Min.* 8:33.

van't Hof, A. E., Edmonds, N., Dalikova, M., Marec, F., and Saccheri, I. J. (2011). Industrial melanism in British peppered moths has a singular and recent mutational origin. *Science* 332, 958–960. doi: 10.1126/science.1203043

Vinuela, A., Snoek, L. B., Riksen, J. A., and Kammenga, J. E. (2010). Genome-wide gene expression regulation as a function of genotype and age in *C. elegans*. *Genome Res.* 20, 929–937. doi: 10.1101/gr.102160.109

Vinuela, A., Snoek, L. B., Riksen, J. A., and Kammenga, J. E. (2012). Aging Uncouples Heritability and Expression-QTL in *Caenorhabditis elegans*. *G3* 2, 597–605. doi: 10.1534/g3.112.002212

Wickham, H. (2009). *Ggplot2 : Elegant Graphics for Data Analysis*. New York, NY: Springer.

Wickham, H. (2018). *tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions*. Available online at: https://mran.microsoft.com/snapshot/2017-02-04/web/packages/tidyr/index.html (accessed May 29, 2020).

Wickham, H. F., Henry, L., and Müller, K. (2018). *dplyr: A Grammar of Data Manipulation*. Available at: dplyr: A Grammar of Data Manipulation (accessed May 29, 2020).

Zahurak, M., Parmigiani, G., Yu, W., Scharpf, R. B., and Berman, D. (2007). Pre-processing Agilent microarray data. *BMC Bioinform.* 8:142. doi: 10.1186/1471-2105-8-142

Zdraljevic, S., Fox, B. W., Strand, C., Panda, O., and Tenjo, F. J. (2019). Natural variation in *C. elegans* arsenic toxicity is explained by differences in branched chain amino acid metabolism. *eLife* 8:e40260.

Zdraljevic, S., Strand, C., Seidel, H. S., Cook, D. E., Doench, J. G., Andersen, E. C., et al. (2017). Natural variation in a single amino acid substitution underlies physiological responses to topoisomerase II poisons. *PLoS Genet.* 13:e1006891. doi: 10.1371/journal.pgen.1006891

Zych, K., Snoek, B. L., Elvin, M., Rodriguez, M., and Van der Velde, K. J. (2017). reGenotyper: Detecting mislabeled samples in genetic data. *PLoS One* 12:e0171324. doi: 10.1371/journal.pone.0171324

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.