More Than the Sum of Its Parts: A Complex Epistatic Network Underlies Natural Variation in Thermal Preference Behavior in *Caenorhabditis elegans*

Bryn E. Gaertner,^{*,†,1} Michelle D. Parmenter,^{*,2} Matthew V. Rockman,^{*} Leonid Kruglyak,[§] and Patrick C. Phillips^{*,3}

*Institute of Ecology and Evolution and [†]Institute of Neuroscience, University of Oregon, Eugene, Oregon 97403, [‡]Center for Genomics and Systems Biology, New York University, New York, New York 10012, and [§]Lewis-Sigler Institute for Integrative Genomics, Department of Ecology and Evolutionary Biology, and Howard Hughes Medical Institute, Princeton University, Princeton, New Jersey 08540

ABSTRACT Behavior is a complex trait that results from interactions among multiple genes and the environment. Both additive and nonadditive effects are expected to contribute to broad-sense heritability of complex phenotypes, although the relative contribution of each of these mechanisms is unknown. Here, we mapped genetic variation in the correlated phenotypes of thermal preference and isothermal dispersion in the nematode *Caenorhabditis elegans*. Genetic variation underlying these traits is characterized by a set of linked quantitative trait loci (QTL) that interact in a complex epistatic network. In particular, two loci located on the X chromosome interact with one another to generate extreme thermophilic behavior and are responsible for ~50% of the total variation. Our results demonstrate that simultaneously considering the influence of a quantitative trait locus (QTL) on multiple scales of behavior can inform the physiological mechanism of the QTL and show that epistasis can explain significant proportions of otherwise unattributed variance within populations.

THE precise nature of the genetic variants underlying complex traits within natural populations has been the subject of a century-old (and often contentious) debate (Provine 1971). It is still unclear whether the genetic component of phenotypic variation tends to be generated predominantly via the contributions of many loci, each with small additive effects (Fisher 1918); a smaller number loci, some with alleles of large effects segregating in a largely Mendelian fashion (Bateson 1913); and/or loci whose allelic effects are strongly dependent on the effects of alleles at other loci [epistasis (Wright 1932)]. Ultimately, the genetic basis of quantitative variation will almost certainly depend on the functional and evolutionary processes that have served to shape the variation over time (Hanson 2006). In the case of gene interactions, complexities in the underlying genetic architecture generated by epistasis can obscure the genotype-phenotype relationship, especially from a statistical point of view (Phillips 1998, 2008; Cordell 2002; Zuk et al. 2012). When a specific combination of alleles at multiple loci is required to produce a particular phenotype, then any one allele will frequently fail to significantly associate with variation in that phenotype, even though such loci can play an important role in the overall pattern of genetic variation within the population (Whitlock et al. 1995). In general, epistasis will reveal itself as a nonadditive interaction between the effects of two independent genes and is expected to emerge at some level from any complex regulatory system underlying a particular phenotype (Dixon et al. 2009). Although there is a general impression tracing back to R. A. Fisher that epistasis is of little importance for understanding genetic variation (Hill et al. 2008; Crow 2010), the reality is that we still actually know very little about the incidence or significance of gene interactions for structuring

Copyright © 2012 by the Genetics Society of America

doi: 10.1534/genetics.112.142877

Manuscript received June 12, 2012; accepted for publication October 9, 2012 Supporting information is available online at http://www.genetics.org/lookup/suppl/ doi:10.1534/genetics.112.142877/-/DC1.

¹Present address: Department of Genetics, North Carolina State University, Raleigh, NC 27565.

²Present address: Department of Genetics, University of Wisconsin, Madison, WI 53706.

³Corresponding author: Institute of Ecology and Evolution, 318 Pacific Hall, 5289 University of Oregon, Eugene, OR 97403-5289. E-mail: pphil@uoregon.edu

the genetics and evolution of complex phenotypes within natural populations (Phillips 2008).

There is perhaps no more complex a phenotype than behavior, as it integrates across environmental stimuli, neural networks, physiology, and cellular function. For example, a behavior such as the response to temperature flux requires that an organism sense the temperature, assess the fitness and performance impacts of that temperature, and then migrate accordingly. Thermal preference is a particularly informative trait because of the direct tie between behavior and fitness, since migration to a thermal optimum influences growth rate and physiological stress (Huey 1974; Huey and Berrigan 2001). Studying thermal preference in a quantitative genetic framework thus presents an opportunity to not only understand the complex genetic architecture of behavior, but also place its variation in an appropriate evolutionary context.

Thermal preference, as assayed experimentally by measuring migration on a thermal gradient, varies among wild isolates of the nematode Caenorhabditis elegans (Anderson et al. 2011). The neural and molecular bases of this behavior are well characterized in the canonical laboratory strain of this model organism, with thermosensation and thermotaxis under the control of only a few neurons and a well-defined class of genes defined via mutagenesis (Garrity et al. 2010). Fitness as measured by both intrinsic rate of increase and total reproductive output is correlated with thermal preference (Anderson et al. 2011), but the genetic architecture of this variation is unknown. In this study, we used a set of recombinant inbred advanced intercross lines (RIAILs) in C. elegans (Rockman and Kruglyak 2009) to identify both additive and nonadditive genetic components to the phenotypes associated with thermal preference, finding that the majority of natural genetic variation for this trait is determined by highly complex interactions among multiple loci.

Materials and Methods

Worm maintenance methods

Strains were stored at -80° and were thawed and maintained in standard conditions (Brenner 1974) for 1 month prior to phenotyping. Lines were maintained on standard 10-cm NGM-lite petri dishes seeded with *Escherichia coli* OP-50 (*Caenorhabditis* Genetics Center, University of Minnesota, Minneapolis) at 20° and chunked twice weekly. The parental strains used in this work were N2, originally isolated from Bristol, England, but long maintained in the laboratory, and CB4856 (abbreviated HW), originally isolated from Hawaii (Wicks *et al.* 2001).

RIAIL construction

RIAILs were generated as described in Rockman and Kruglyak (2008, 2009). Briefly, N2 and HW were reciprocally crossed in P_0 , F_1 , and F_2 generations. F_3 progeny were random pair mated for 10 generations and then inbred by pick-

ing individual hermaphrodites for 10 generations. Lines were genotyped at 1455 single-nucleotide polymorphisms.

Nearly Isogenic Line construction

Nearly isogenic lines (NILs) were generated in both N2 and HW genetic backgrounds, using marker-assisted selection. Restriction-fragment length polymorphic markers (RFLPs) were selected as described originally in Wicks et al. (2001). Candidate RIAILs were selected based on their genotypes at the QTL peaks. These RIAILs were then crossed and backcrossed into the background contragenic to the genotype across the QTL for at least 10 generations. At each generation of backcrossing, individuals were genotyped across 10 evenly spaced markers on the X chromosome to identify recombination events. Double-NILs were constructed by crossing appropriate NILs to each other and genotyping F_3 progeny at loci relevant for identifying meiotic recombination. This was used to generate the following strains: PX513 [thrm-3(fxIR1), an HW introgression at thrm-3 in an otherwise N2 background], PX514 [thrm-5(fxIR2), an HW introgression at thrm-5 in an N2 background], PX515 [thrm-3(fxIR1);thrm-5(fxIR2)], PX516 [thrm-3(fxIR3), an N2 introgression at *thrm-3* in an otherwise HW background], PX517 [thrm-5(fxIR4), an N2 introgression at thrm-5 in an HW background], and PX518 [thrm-3(fxIR3);thrm-5 (fxIR4)].

Thermal gradient construction and calibration

Three replicate linear thermal gradients each with a slope of $\sim 1.0^{\circ}$ /cm were used in all assays and constructed as described previously (Anderson et al. 2007). For each gradient an aluminum slab was placed over two separate piping systems, one carrying heated (55°) and one carrying cooled (4°) liquid. A 10×17 -cm slab of NGM-lite agar was placed on the aluminum slab and enclosed in a plexiglass frame. The slope of each of the three gradients was determined by assessing their temperature at 1-cm intervals and fitting these readings by linear regression. In all cases, $r^2 > 0.98$, indicating that the gradients were linear. The slopes of the gradient remained stable over repeated measurements but were slightly different among the three replicates. This slight difference in slope (ranging from 0.8°/cm to 1.1°/cm) did not affect the measured thermal preference within a line.

Thermal preference and isothermal dispersion assay

Lines were age synchronized by sodium hypochlorite treatment (Stiernagel 1999). Approximately 600 isogenic L1's were plated onto three 6-cm NGM-agar plates seeded with *E. coli* OP-50 for a total of 1800 individuals. These individuals were then raised at 20° for 48 hr until reaching the L4 larval stage. Worms were washed from each plate with 1 mL S Basal into 1.5-mL microcentrifuge tubes. Supernatant S Basal was pipetted off once the worms settled. Worms, suspended in ~100 μ L of S Basal, were transferred via filter paper onto the gradient-equilibrated NGM-agar gel and placed equidistant between either edge of the gel at 24°. Approximately half of the worms were lost in this transfer such that \sim 250 individuals per gel were assayed. To ensure that thermal preference, and not any behavioral response to starvation, was being assayed, agar gels were seeded with *E. coli* (OP-50) 24 hr prior to use. Worms were permitted to travel freely across the gel for 1 hr. As worms on food move between 0.025 mm/sec and 0.25 mm/sec (Arous *et al.* 2009), there is a potential total movement of between 9 and 108 cm over the duration of the experiment. After 1 hr, individual positions were scored by overlaying transparency film and marking worm locations with a semipermanent marker. In all, the responses of an average of 700 individuals per RIAIL line and 91,831 individuals overall were measured using this approach.

Statistical analysis of phenotypes

Once scored, transparencies were scanned and analyzed in ImagePro Plus 5.1 (Media Cybernetics, Bethesda, MD). The Cartesian coordinates of each individual worm's marked position were recorded and related back to the calculated slope of the gradient such that the temperature of each individual (derived from the "y" Cartesian coordinate) as well as the distance it moved laterally (derived from the "x" Cartesian coordinate) from where it was initially placed was identified. These two values represent the thermal preference and isothermal dispersion of any individual. Overall RIAIL and NIL phenotypes were found by performing a nested randomeffects ANOVA, with a main effect of line and a nested effect of replicate using JMP 9.0 (SAS Institute, Cary, NC). The average RIAIL response was determined using the leastsquares mean estimated from this analysis, although using individual line means yielded essentially identical results. Thermal preference and isothermal dispersion served as the two principal phenotypes to be mapped. A third phenotype, which found thermal preference after taking movement into account, was found by using the residuals of regressing individual thermal preference onto isothermal dispersion.

QTL mapping of phenotypes

Interval mapping and two-dimensional genome scans were performed as implemented in R/QTL (Broman *et al.* 2003) using genotypes and phenotypes provided in Supporting Information, File S1. Due to the tight linkage of identified QTL, there was not sufficient power to perform multiple-interval mapping. For interval mapping, genotype probabilities were calculated using Haldane's map function with a 1-cM step size and an error probability of $1-10^{-4}$. Whole-genome scans with a single-QTL model for isothermal dispersion and thermal preference were performed using the EM algorithm (Lander and Botstein 1989), although results were qualitatively the same when Haley–Knott regression was used instead (Haley and Knott 1992). When appropriate, covariate markers were used to account for variation due to previously identified large-effect QTL (see *Results*) in the single-QTL model. The analysis was permuted 1000 times to obtain a genome-wide or a chromosome-wide LOD significance threshold (Churchill and Doerge 1994). For single-QTL models, a LOD score of \sim 3.19 signified a genome-wide *P* < 0.05 threshold.

A two-dimensional genome scan was used to test for the likelihood of interacting QTL for all pairwise combinations of intervals by including a statistical interaction effect between intervals (Dupuis *et al.* 1995), which is a signature of epistasis. This scan was performed using the EM algorithm (Lander and Botstein 1989) with no specified covariates and a 1-cM step size. Genome-wide significance levels for the reduced (epistasis only) and full models were determined using a permutation test. LOD scores for the full model (L_f) at locus pair *j*, considering the additive and interactive effects of any two intervals (q_1 , q_2), are calculated by the equation

$$LOD_j = \log_{10} L_f(q_1, q_2) - \log_{10} L_0,$$

where L_0 is the null model. Epistasis LOD scores are calculated as

$$LOD_i = \log_{10} L_f(q_1, q_2) - \log_{10} L_a(q_1, q_2),$$

where L_a is the likelihood under the additive model. The difference between these two likelihoods demonstrates the likelihood of the interaction term by itself (Broman *et al.* 2003).

Results

Thermal preference and isothermal dispersion vary continuously among RIAILs

Consistent with previous results (Anderson et al. 2007; Jurado et al. 2009), we find that when started at a relatively warm position, the N2 parental strain remained in the warm portion of the gradient (Figure 1A) ($\bar{X} = 21.87^{\circ}$, SEM = 0.32°) while the HW strain migrated to a much colder region of the gradient ($\bar{X} = 17.62^{\circ}$, SEM = 0.28°). In contrast, thermal preference among the RIAILs displayed substantial transgressive segregation, ranging from 17.34° to 27.01°, with approximately half of all lines assayed having a thermal preference that was warmer than that of the more thermophilic N2 parent. The warmer thermal preferences are well above those observed for other natural isolates (Anderson et al. 2011) and are also above the upper thermal limit for continuous population growth (Anderson et al. 2011). Almost half (49.4%) of the total variance in thermal preference was attributable to between-line or genetic differences $(F_{118, 247} = 22.01, P < 0.001).$

Isothermal dispersion for the parents ranged from 0.68 cm/hr (N2) to 2.13 cm/hr (HW), while that for the RIAILs ranged from 0.44 cm/hr to 2.86 cm/hr (Figure 1B). Approximately 16% of total variance in isothermal dispersion was explained by among-line differences ($F_{118, 247} = 15.23$, P < 0.001). Unlike thermal preference, then, isothermal dispersion



Figure 1 Variation in isothermal dispersion and thermal preference. Each bar represents the lines mean from a recombinant inbred advanced intercross line between a *C. elegans* parental line derived from either Bristol, England (N2) or Hawaii (CB4856, "HW"). RIAILs are rank ordered for temperature (A) and distance (B). Dashed and dotted horizontal lines represent HW and N2 phenotypes, respectively.

was well bounded by the parental values. There was a slight but significant correlation between isothermal dispersion and thermal preference among lines ($r^2 = 0.20$, P < 0.01; Figure S1).

For both thermal preference and isothermal dispersion there were some differences among replicates (isothermal dispersion, $F_{246,85078} = 9.64$, P < 0.001; thermal preference, $F_{246,85079} = 36.44$, P < 0.001), although these differences did not display any systematic bias and accounted for only 3.1% and 6.8% of total phenotypic variance, respectively.

Several loci contribute nonadditively across phenotypes to explain variation in thermal preference

To identify the loci responsible for variation in isothermal dispersion and thermal preference, we used interval mapping with and without covariates. For isothermal dispersion, we found evidence for a single locus of large effect on the X chromosome (Figure 2A, LOD = 20, P < 0.0001). This peak is centered over *npr-1*, allelic variation at which has been identified as responsible for variation in a variety of behavioral responses, including the tendency to "clump" on a petri dish (de Bono and Bargmann 1998), the response to thermal stress (Glauser et al. 2011), avoidance of high oxygen concentrations (McGrath et al. 2009; Persson et al. 2009), and tendency to leave food patches (Bendesky et al. 2011). Because the N2 npr-1 allele was previously identified as laboratory derived and was shown to interact epistatically with other loci (McGrath et al. 2009), we accounted for variation at this locus in subsequent isothermal migration analyses by using the marker most closely associated with this peak as an interactive covariate. In doing so, we found that the original QTL on the X chromosome is likely to be composed of at least two QTL, with a significant second QTL emerging \sim 800 kb distal to *npr-1* (Figure 2B, LOD = 4.95,

P < 0.001). We named this second locus *disp-1* (*DISPersion*). To identify the nature of the relationship between the *npr-1* and *disp-1* loci, we tested for interaction effects between the two loci (Figure 3). We found that lines possessing N2 alleles at both loci tend to disperse only slightly (1.02 ± 0.043 cm/hr), whereas lines possessing one HW allele at either *disp-1* or *npr-1* moved significantly farther (1.86 ± 0.139 cm/hr, $F_{1,110} = 95.95$, P < 0.0001) (Figure 3). Dispersal of a subset of the RIAILs was assayed on an isothermal (20°) agar slab in the absence of a gradient and yielded qualitatively similar mapping results (Figure S2).

For thermal preference, we identified two significant and one strongly suggestive main-effect QTL, all of which were located on the X chromosome (Figure 2, A and C). We named these THeRMal preference (thrm) loci in proximaldistal order on the chromosome: thrm-1, thrm-2, and thrm-4. thrm-2 colocalized with the disp-1 locus described above, so further analysis uses *disp-1/thrm-2* to indicate this locus. The same QTL were identified by analysis of residuals of thermal preference regressed on isothermal dispersion, implying that the genetics of thermal preference are largely independent of isothermal dispersion phenotypes. No QTL colocalized with *npr-1*, and incorporation of a marker at *npr-*1 as an interactive covariate did not substantially change the LOD profile for thermal preference, implying that the QTL mapped for this phenotype are also largely independent of npr-1. No genes known to cause thermotaxis defects when mutated colocalized within the 1.5-LOD confidence intervals of these novel QTL, suggesting novel variants are contributing to observed thermal preference differences (Table S1).

Accounting for variation at *disp-1* significantly changed the LOD profile of the QTL peaks in thermal preference (in contrast to using *npr-1* as a covariate). To explore this further, we explicitly tested for interactions between genotypes at *disp-1/thrm-2* and genotypes at *thrm-1* and *thrm-4* (Figure 4).



Map position on X Chromosome (cM) Genomic location on X Chromosome (MBp)

Here, an interesting pattern emerged. We found that for lines with an N2 genotype at *disp-1/thrm-2*, variation in *thrm-1*, but not in *thrm-4* explained variation in thermal preference. In this case, lines with an HW genotype at *thrm-1* had a cooler thermal preference (22.51° ± 0.44°) and lines with an N2 genotype at *thrm-1* had a warmer thermal preference (23.95° ± 0.26°, $F_{1, 108} = 8.06$, P < 0.01). Conversely, we found that for lines with an HW genotype at *disp-1/thrm-2*, variation in *thrm-4*, but not in *thrm-1* explained variation in thermal preference. In this case, lines with an HW genotype at *thrm-4* had a cooler thermal preference (21.53° ± 0.33°) and lines with an N2 genotype at *thrm-4* had a warmer thermal preference (23.31° ± 0.52°, $F_{1, 108} = 8.96$, P < 0.01, summarized in



Figure 3 Epistasis between two QTL for isothermal dispersion. Across lines, two N2 genotypes (at both *npr-1* and *disp-1*, schematized as red boxes for both loci) result in significantly reduced isothermal dispersion. Possessing an HW allele (blue) at either *npr-1* or *disp-1* masks the effect of an N2 allele at the other locus, resulting in a high-movement phenotype. Error bars are ± 2 SEM; ***significant difference at P < 0.0001.

Figure 2 QTL maps of isothermal dispersion and thermal preference. (A) Isothermal dispersion (orange) and thermal preference (purple) map to the X chromosome. Genome-wide P = 0.05 significance threshold = 3.14 LOD (horizontal line). (B) Isothermal dispersion on the X chromosome. The original QTL peak (dashed orange line) is at \sim 4.7 Mb. When accounting for variation at this locus (npr-1, represented with an arrow), the peak goes to zero as it is included as a covariate in the subsequent map (solid line). In this analysis, a second significant QTL appears at \sim 5.6 Mb (*disp-1*). (C) Thermal preference maps to three significant loci on the X chromosome. thrm-1, thrm-2, and thrm-4 significantly contribute additively to variation in thermal preference (black arrows). thrm-3 and thrm-5 (gray arrows) are not significant by themselves, but when considered interactively contribute significant proportions of variance to thermal preference. QTL regions are schematized for each (B) isothermal dispersion and (C) thermal preference below the x-axis, with light gray boxes indicating regions that explain a significant proportion of genetic variation.

Table S2). In summary, we found that thermal preference variation among RIAILs was best explained by accounting for genetic variation in the isothermal dispersion QTL of *disp-1/thrm-2*, in that the genotype at *thrm-4* explained variation in thermal preference for lines with an HW genotype at *disp-1/thrm-2*, but the genotype at *thrm-1* explained variation in thermal preference for lines with an N2 genotype at *disp-1/thrm-2* (Figure 8).

Given the preponderance of nonadditive interactions among significant QTL, we performed an unbiased test of all pairwise combinations of 1-cM intervals across the genetic map to detect epistasis for both isothermal dispersion and thermal preference (Figure S3, X chromosome, and Figure 5, A and B) to determine whether suggestive but not significant peaks were accounting for nonadditive genetic variation. We found no such loci for isothermal dispersion. However, for thermal preference we identified strong evidence for epistasis between two loci, also on the X chromosome, termed thrm-3 and thrm-5 (Figure 2, A and C, and Figure 5B, LOD = 4.62, P < 0.001). On their own, *thrm-3* and *thrm-5* did not display significant main effects ($t_{118} = -1.15$, 1.19, respectively, for thrm-3 and thrm-5, P > 0.1; see also Figure 2C). However, when considered together, the interaction effect between thrm-3 and thrm-5 explained half the overall genetic variation in thermal preference, accounting for virtually all of the transgressive segregation observed in the RIAILs. Averaging across effects at all other loci, we found that RIAILs with a "matching" genotype between thrm-3 and thrm-5 (i.e., either N2 genotype at both loci or HW at both loci) exhibited thermal preferences consistent with the phenotypes of the parents. However, "mismatching" genotypes across loci (i.e., N2 genotype at one locus and HW at the other) corresponded with the excessively thermophilic phenotype (Figure 6).



Genetic mismatch between two loci causes extraordinarily thermophilic behavior

To confirm the genetic basis for the identified statistical epistasis pattern between thrm-3 and thrm-5, we generated NILs, using marker-assisted selection. NILs were made in both genetic backgrounds, and these lines were then run in triplicate on linear gradients. We found that the thermophilic interaction effect inferred in the RIAILs was verified in the NILs. Regardless of overall genetic background, our evidence for epistatic interactions remained because singleintrogression NILs preferred warm temperatures ($F_{1,10}$ = 194.75, P < 0.0001; Figure 7). Additionally, the doubleintrogression NIL in the N2 genetic background has a thermal preference that is not significantly different from that of the N2 parent $(22.05^{\circ} \pm 0.38^{\circ}, 21.07^{\circ} \pm 0.27^{\circ}, \text{ respec-}$ tively), and the double-introgression NIL in the HW genetic background has a thermal preference that is significantly colder than that of the single-introgression NILs, although still slightly more thermophilic than that of the HW parent $(19.49^{\circ} \pm 0.39^{\circ}, 17.67^{\circ} \pm 0.22^{\circ}$ respectively; Figure 7). Put simply, a genetic mismatch between these two loci in either genetic background causes excessively thermophilic behavior, and a genetic match between these loci permits the main-effect QTL (thrm-1, disp-1/thrm-2, and thrm-4) to determine thermal preference (Figure 8).

Discussion

Complex epistasis underlying thermal preference behavior

Nearly all ecologically and evolutionary relevant traits are complex, as are most common human diseases such as heart disease, autism, and diabetes (Manolio *et al.* 2009; Gibson 2010). Yet despite tremendous effort, success in identifying the genetic basis of complex traits has been achieved only in a handful of cases, usually when the underlying variation behaves in a largely Mendelian fashion (Rockman 2012). One possible explanation for this lack of success is that most studies still lack sufficient statistical power, especially in terms of detecting rare alleles with moderate effects. However, another possible explanation is that nonadditive genetic interactions (epistasis) are responsible for much of **Figure 4** Phenotypic epistasis exists between *disp-1* and thermal preference QTL. Between high-moving and low-moving strains, different thermal preference loci explain variance in thermal preference, shown here using the marginal mean thermal preference of each genotype at each locus. (A) Genotype at *thrm-1* explains variation in thermal preference only for lines with an N2 (red) genotype at *disp-1/thrm-2*; for lines with an HW (blue) genotype at *disp-1/thrm-2*, there is no difference in thermal preference due to genotype at *thrm-1*. (B) Genotype at *thrm-4* explains variation in thermal preference only for lines with an N2 genotype at *disp-1/thrm-2*; for lines with an N2 genotyp

the segregating variation and do so in a fashion that obscures the effects of individual loci (Phillips 2008; Moore and Williams 2009). This is because in natural populations, such interactions depend on the generation of particular combinations of alleles that, because of their infrequent co-occurrence, are ephemeral and may be difficult to identify using association-mapping approaches. Such genetic interactions are instead best identified using experimental mapping crosses and a substitution-based approach. Here, we used these approaches to characterize the genetic architecture underlying thermal preference in *C. elegans*, finding it to be highly complex, involving multiple layers of epistasis. In particular, \sim 50% of broad-sense heritable variation is caused by a compensatory relationship between two loci.

Because of limited power to investigate all pairwise interactions, a typical experimental approach in mapping studies is to first identify main-effect QTL and then test for interactions among them (Culverhouse et al. 2002). This approach, although utilized in part of this study, suffers from ascertainment bias, particularly because loci with the largest interaction effects are precisely those that are likely to have smaller relative individual effects (Phillips 2008). For thermal preference, a comprehensive pairwise genome scan identified statistical epistasis (Fisher 1918) in the form of two loci that interact with each other to generate extreme thermophilic behavior, a phenotype observed in our mapping population but whose variance was not explained through main-effect QTL. The asymmetric transgressive segregation we observed among lines is a signature of nonadditive variation (Rieseberg et al. 1999). We used single and double NILs in both genetic backgrounds at both interactive loci to demonstrate unequivocally that nonadditive interactions between these two loci cause the excessively thermophilic behavior observed in the RIAILs, well beyond the range of thermal preference observed in other natural isolates (Anderson et al. 2011). Especially telling in this case was the fact that double-introgression NILs were able to recover nearly all of the original parental phenotype that had been strongly disrupted by single substitution of either one of the other loci. This strategy of using introgression lines to functionally test these interacting QTL allowed us to connect Fisher's original concept of statistical epistasis estimated from variation segregating in a population with



Figure 5 Pairwise QTL scan for significant interaction effects. Two loci on the X chromosome, at \sim 122 and 206 map units and colocalizing precisely with *thrm-3* and *thrm-5* from the interval mapping (Figure 2C), show a strong interaction effect.

the more traditional, Batesian concept of compositional epistasis estimated via the segregational effects of specific loci (Phillips 2008). Although we cannot say at this point precisely how (or whether) the alleles underlying *thrm-3* and *thrm-5* are functionally interacting [the final common usage of "epistasis" (Phillips 2008)], we can say that in two separate genetic backgrounds (HW and N2), an allelic mismatch causes thermophilic behavior. By taking advantage of known mutations in these regions, in addition to the well-studied developmental and physiological characteristics of the thermotaxis neural circuitry, we should be able to identify not just the novel genes involved in this thermophilic behavior but also the precise nature of their epistatic interactions.

Phenotypic epistasis is generated by functional hierarchies

Breaking the linkage between the *thrm-3* and *thrm-5* loci, as seen in our single-introgression NILs, masks the ability of main-effect QTL to influence thermal preference. By "rescuing" this linkage through double NILs, which are matching at *thrm-3* and *thrm-5* but in the contragenic background, we were able to show that the *thrm-3/thrm-5* complex is itself upstream of, or epistatic to, the main-effect QTL modules we identified using interval mapping (Figure 8) and that in the presence of matching genotypes at *thrm-3* and *thrm-5*, the main-effect QTL we identified predicted thermal preference. The nature of these interactions suggests a hierarchical epistatic network underlying natural variation in thermal preference (Figure 8).

The downstream module of this network is itself generated by interactions between loci related to hierarchically structured phenotypes of movement and thermal preference, which created a pattern of epistasis similar to that expected for directly interacting gene products. Under this



Figure 6 Epistatic interaction of *thrm-3* and *thrm-5*. Lines with an N2 genotype at both *thrm-3* and *thrm-5* have a thermal preference approximately consistent with that of the parental N2 line. Lines with an HW genotype at both *thrm-3* and *thrm-5* have a thermal preference approximately consistent with that of the parental HW line. Discrepancies are due to averaging across other loci contributing to thermal preference variation directly. Allelic heterogeneity between loci (HW at one locus and N2 at another) predicts thermophilic behavior. Error bars are ±2 SEM, and homogenous subsets [Tukey's honestly significant differences (HSD), *P* < 0.01] are denoted with lowercase letters.

"phenotypic epistasis" (Segrè et al. 2004), the expression of one trait is influenced by the expression of other traits. Specifically, we found that worms can be divided into high- and low-movement groups, where worms with an N2 genotype at disp-1 tend to move little, conditional on the N2 npr-1 genotype, and worms with an HW genotype at *disp-1* tend to move substantially more. This is not due to a deficiency in the assay in which, say, expression of specific thermal preference phenotypes becomes contingent upon movement per se. Instead, it is clear that even low-moving lines are "motivated" to go to colder temperatures: some lines with an average isothermal dispersion of ~ 1 cm/hr still migrate $>4^{\circ}$ cooler than the start temperature (Figure S1), a distance of \sim 4 cm. Thus, despite a tendency to not disperse in isothermal conditions, these lines are still capable of navigating a thermal gradient and do so to find colder temperatures. Additionally, lines predicted to be thermophilic because of a mismatching genotype between thrm-3 and thrm-5 migrate to warm temperatures, despite also possessing an N2 "low-movement" phenotype. This is true for both RIAIL and NIL genotypes.

Thus movement *per se*, as measured by total distance traveled, cannot be the sole explanation for the phenotypic epistasis we observed. Extensive behavioral and neurophysiological analysis of the N2 strain navigating on a chemical gradient (Pierce-Shimomura *et al.* 1999; Iino and Yoshida 2009; McCormick *et al.* 2011) has shown that individuals use different search strategies as they experience different concentration changes (*i.e.*, klinotaxis *vs.* klinokinesis), with additional evidence that N2 uses the same strategies in navigating a thermal gradient (Ryu and Samuel 2002) or thermal step (Zariwala *et al.* 2003; McCormick *et al.* 2011). In this context, rate of movement affects the effective concentration



Figure 7 Thermal preference of nearly isogenic lines (NILs) with introgressions at *thrm-3*, at *thrm-5*, or at both. Single NILs are significantly more thermophilic than the parent with the same genetic background. Double NILs have a thermal preference not significantly different from that of the parents with the same overall genetic background. Error bars are ± 2 SEM, and homogenous subsets (Tukey's HSD, P < 0.05) are denoted with lowercase letters.

change that each individual is experiencing. This presents a connection between the movement-conditional QTL presented in this study and the different navigation strategies: perhaps the *thrm-1* and *thrm-4* QTL have different effects in high- and low-moving individuals because fundamentally different navigation strategies are being used.

The unusual genomic history of C. elegans

In contrast to the complex genetic interactions seen in thermal preference, isothermal dispersion displayed a more classic pattern of epistasis in which the low-movement phenotype is generated only when an individual possesses N2 alleles at both causal loci. Interestingly, these loci are located in what was initially identified as a single largeeffect OTL, with one of the interacting loci directly centered over the previously identified laboratory-derived allele of *npr-1*. Considering the hundreds of generations that the canonical N2 strain has lived under laboratory conditions, it is not surprising that fixation due to drift, as well as selection on mutations beneficial for a laboratory environment, has occurred (McGrath et al. 2009, 2011a; Weber et al. 2010). The latter is clearly the case for *npr-1*, in which the laboratoryderived mutation simultaneously inhibits burrowing in agar and clumping on a plate, two characteristics that would be selected against under laboratory rearing conditions and facilitated by increasing tolerance of atmospheric levels of oxygen in a nematode that appears to spend most of its natural life in rotting substrates (Kiontke et al. 2011). The laboratory-derived nature of this allele is apparent both through association mapping across wild isolates (McGrath et al. 2009) and through whole-genome resequencing of N2 against its close relatives [LSJ1 and LSJ2 (Weber et al. 2010; McGrath et al. 2011b)] and has been implicated as a causal variant in many behaviors that differ between N2 and HW (Gaertner and Phillips 2010). In this study, npr-1 was clearly



Figure 8 A hierarchical epistatic network determining thermal preference. Genotypes between *thrm-3* and *thrm-5* (top, boxed) must be "matching" for loci in other regions of the genome to influence thermal preference (subsequent levels of the hierarchy). In this case, for lines with an N2 genotype at *disp-1/thrm-2*, variation at *thrm-1* explains variation in thermal preference but for lines with an HW genotype at *disp-1/thrm-2*, variation at *thrm-4* explains variation in thermal preference. Lines with a "mismatch" genotype between *thrm-3* and *thrm-5*, regardless of genotype across the genome, display thermal preference that is excessively thermophilic.

implicated in variation in isothermal dispersion. However, the effect of this allele is masked in the presence of an HW genotype at *disp-1* (see also McGrath *et al.* 2009). Additionally, for thermal preference, genetic variation at *npr-1* did not change the proportion of variance explained at any other locus, strongly suggesting that variation in the behavior of thermal preference is not due to *npr-1*–derived laboratory adaptation.

The finding of pervasive epistasis, especially among a large number of tightly linked loci, may be more common in C. elegans than in other animals because of their rare androdieocious mating system consisting of self-fertile hermaphrodites and the occasional outcrossing male. Pervasive self-fertilization tends to decrease the effective recombination rate and greatly increases the probability that interacting alleles will be found (and evolve) in the same genetic background (Fenster et al. 1997; Nordborg 2000). In part due to this mating system, most variation in genetic diversity and gene expression in C. elegans can be explained by genetic hitchhiking (Graustein et al. 2002; Cutter and Payseur 2003; Rockman et al. 2010; Andersen et al. 2012), with reduced genetic variation locked in long haplotype blocks and a strong negative correlation between polymorphism frequency and gene density across the chromosome. In this context, interacting loci with effects that balance one another may evolve in regions of low recombination, and breaking these regions apart through generations of mapping crosses can act to decanalize the assayed phenotype (Gibson 2009).

There is a noted mismatch between the estimated outcrossing frequency in nature based on genomic estimates of global effective population size (Rockman and Kruglyak 2009) and empirical observation of heterozygosity and male frequency (Barrière and Félix 2007) in locally isolated populations of C. elegans (Anderson et al. 2010). This incongruence supports a model of selection against incompatible genotype combinations such as the ones described here [and elsewhere (Seidel et al. 2008; 2011)], where local haplotype and phenotypic diversity might be generated through decanalizing epistasis but masked in a global survey (Barrière and Félix 2007; Anderson et al. 2010). Interestingly, this pattern of reduced genetic diversity does not appear to extend to the X chromosome, which displays slightly more nucleotide diversity and lower and more even linkage disequilibrium (Andersen et al. 2012) and was where all of our identified thermal preference and isothermal dispersion QTL are located. Thus even in regions with a more uniform distribution of polymorphism, recombination, and haplotype diversity, epistasis can persist, suggesting a more central role of these nonadditive interactions in generating diversity within natural populations.

In summary, our study demonstrates the central role of epistasis in generating natural variation in behavior. Epistasis was identified statistically through mapping and connected to compositional epistasis through the generation of introgression lines in two genetic backgrounds. By generating these lines, we were able to demonstrate that interactions also exist between complexes of epistatic loci and hierarchically across phenotypes. Thus, understanding how to construct the genotype–phenotype map through a cellular, physiological, and organismal context requires not only statistical, but also experimental analysis of these interactions.

Acknowledgments

We thank W. Cresko, J. Fierst, R. Reynolds, and two anonymous reviewers for critical reading of the manuscript and S. Lockery, M. Weir, T. Herman, and the Phillips Laboratory for helpful discussions. This work was supported by grants IOS-0909816 (to B.E.G. and P.C.P.) and DEB-1120417 (to P.C.P.) from the U.S. National Science Foundation and grants R01-HG004321 (to L.K.) and R01-GM089972 (to M.V.R.) from the National Institutes of Health.

Literature Cited

- Andersen, E. C., J. P. Gerke, J. A. Shapiro, J. R. Crissman, R. Ghosh et al., 2012 Chromosome-scale selective sweeps shape Caenorhabditis elegans genomic diversity. Nat. Genet. 44: 285–290.
- Anderson, J. L., L. Albergotti, S. Proulx, C. Peden, R. B. Huey et al., 2007 Thermal preference of *Caenorhabditis elegans*: a null model and empirical tests. J. Exp. Biol. 210: 3107–3116.
- Anderson, J. L., L. T. Morran, and P. C. Phillips, 2010 Outcrossing and the maintenance of males within *C. elegans* populations. J. Hered. 101: S62–S74.
- Anderson, J. L., L. Albergotti, B. Ellebracht, R. B. Huey, and P. C. Phillips, 2011 Does thermoregulatory behavior maximize reproductive fitness of natural isolates of *Caenorhabditis elegans*? BMC Evol. Biol. 11: 157.

- Arous, J. B., S. Laffont, and D. Chatenay, 2009 Molecular and sensory basis of a food related two-state behavior in *C. elegans*. PLoS ONE 4: e7584.
- Barrière, A., and M.-A. Félix, 2007 Temporal dynamics and linkage disequilibrium in natural *Caenorhabditis elegans* populations. Genetics 176: 999–1011.
- Bateson, W., 1913 *Mendel's Principles of Heredity.* Cambridge University Press, Cambridge, UK.
- Bendesky, A., M. Tsunozaki, M. V. Rockman, L. Kruglyak, and C. I. Bargmann, 2011 Catecholamine receptor polymorphisms affect decision-making in *C. elegans*. Nature 472: 313–318.
- Brenner, S., 1974 The genetics of *Caenorhabditis elegans*. Genetics 77: 71–94.
- Broman, K., H. Wu, and Ś. Sen, 2003 R/qtl:QTL mapping in experimental crosses. Bioinformatics 19: 889–890.
- Churchill, G. A., and R. W. Doerge, 1994 Empirical threshold values for quantitative trait mapping. Genetics 138: 963–971.
- Cordell, H. J., 2002 Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. Hum. Mol. Genet. 11: 2463–2468.
- Crow, J. F., 2010 On epistasis: why it is unimportant in polygenic directional selection. Philos. Trans. R. Soc. B. Biol. Sci. 365: 1241–1244.
- Culverhouse, R., B. K. Suarez, J. Lin, and T. Reich, 2002 A perspective on epistasis: limits of models displaying no main effect. Am. J. Hum. Genet. 70: 461–471.
- Cutter, A. D., and B. A. Payseur, 2003 Selection at linked sites in the partial selfer *Caenorhabditis elegans*. Mol. Biol. Evol. 20: 665–673.
- de Bono, M., and C. I. Bargmann, 1998 Natural variation in a neuropeptide Y receptor homolog modifies social behavior and food response in *C. elegans*. Cell 94: 679–689.
- Dixon, S. J., M. Costanzo, A. Baryshnikova, B. Andrews, and C. Boone, 2009 Systematic mapping of genetic interaction networks. Annu. Rev. Genet. 43: 601–625.
- Dupuis, J., P. O. Brown, and D. Siegmund, 1995 Statistical methods for linkage analysis of complex traits from high-resolution maps of identity by descent. Genetics 140: 843–856.
- Fenster, C. B., L. F. Galloway, and L. Chao, 1997 Epistasis and its consequences for the evolution of natural populations. Trends Ecol. Evol. 12: 282–286.
- Fisher, R. A., 1918 The correlation between relatives on the supposition of Mendelian inheritance. Trans. R. Soc. Edinb. 52: 399–433.
- Gaertner, B. E., and P. C. Phillips, 2010 *Caenorhabditis elegans* as a platform for molecular quantitative genetics and the systems biology of natural variation. Genet. Res. 92: 331–348.
- Garrity, P. A., M. B. Goodman, A. D. Samuel, and P. Sengupta, 2010 Running hot and cold: behavioral strategies, neural circuits, and the molecular machinery for thermotaxis in *C. elegans* and *Drosophila*. Genes Dev. 24: 2365–2382.
- Gibson, G., 2009 Decanalization and the origin of complex disease. Nat. Rev. Genet. 10: 134–140.
- Gibson, G., 2010 Hints of hidden heritability in GWAS. Nat. Genet. 42: 558–560.
- Glauser, D. A., W. C. Chen, R. Agin, B. L. MacInnis, A. B. Hellman 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.
- Graustein, A., J. M. Gaspar, J. R. Walters, and M. F. Palopoli, 2002 Levels of DNA polymorphism vary with mating system in the nematode genus *Caenorhabditis*. Genetics 161: 99–107.
- Haley, C. S., and S. A. Knott, 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69: 315–324.
- Hanson, T. F., 2006 The evolution of genetic architecture. Annu. Rev. Ecol. Evol. Syst. 37: 123–157.

- Hill, W. G., M. E. Goddard, and P. M. Visscher, 2008 Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet. 4: e1000008.
- Huey, R. B., 1974 Behavioral thermoregulation in lizards: importance of associated costs. Science 184: 1001–1003.
- Huey, R. B., and D. Berrigan, 2001 Temperature, demography, and ectotherm fitness. Am. Nat. 158: 204–210.
- Iino, Y., and K. Yoshida, 2009 Parallel use of two behavioral mechanisms for chemotaxis in *Caenorhabditis elegans*. J. Neurosci. 29: 5370–5380.
- Jurado, P., E. Kodama, Y. Tanizawa, and I. Mori, 2009 Distinct thermal migration behaviors in response to different thermal gradients in *Caenorhabditis elegans*. Genes Brain Behav. 9: 120–127.
- Kiontke, K. C., M.-A. Felix, M. Ailion, M. V. Rockman, C. Braendle et al., 2011 A phylogeny and molecular barcodes for *Caeno-rhabditis*, with numerous new species from rotting fruits. BMC Evol. Biol. 11: 339–357.
- Lander, E. S., and D. Botstein, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121: 185–199.
- Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff *et al.*, 2009 Finding the missing heritability of complex diseases. Nature 461: 747–753.
- McCormick, K. E., B. E. Gaertner, M. Sottile, P. C. Phillips, and S. R. Lockery, 2011 Microfluidic devices for analysis of spatial orientation behaviors in semi-restrained *Caenorhabditis elegans*. PLoS ONE 6: e25710.
- McGrath, P. T., M. V. Rockman, M. Zimmer, H. Jang, E. Z. Macosko et al., 2009 Quantitative mapping of a digenic behavioral trait implicates globin variation in *C. elegans* sensory behaviors. Neuron 61: 692–699.
- McGrath, P. T., Y. Xu, M. Ailion, J. L. Garrison, R. A. Butcher et al., 2011a Parallel evolution of domesticated *Caenorhabditis* species targets pheromone receptor genes. Nature 477: 321–325.
- McGrath, P. T., Y. Xu, M. Ailion, J. L. Garrison, R. A. Butcher *et al.*, 2011b Parallel evolution of domesticated *Caenorhabditis* species targets pheromone receptor genes. Nature 477: 321–325.
- Moore, J. H., and S. M. Williams, 2009 Epistasis and its implications for personal genetics. Am. J. Hum. Genet. 85: 309–320.
- Nordborg, M., 2000 Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. Genetics 154: 923–929.
- Persson, A., E. Gross, P. Laurent, K. E. Busch, H. Bretes *et al.*, 2009 Natural variation in a neural globin tunes oxygen sensing in wild *Caenorhabditis elegans*. Nature 458: 1030–1033.
- Phillips, P. C., 1998 The language of gene interaction. Genetics 149: 1167–1171.
- Phillips, P. C., 2008 Epistasis–the essential role of gene interactions in the structure and evolution of genetic systems. Nat. Rev. Genet. 9: 855–867.
- Pierce-Shimomura, J. T., T. M. Morse, and S. R. Lockery, 1999 The fundamental role of pirouettes in *Caenorhabditis elegans* chemotaxis. J. Neurosci. 19: 9557–9569.

- Provine, W., 1971 The Origins of Theoretical Population Genetics. University of Chicago Press, Chicago.
- Rieseberg, L. H., M. A. Archer, and R. K. Wayne, 1999 Transgressive segregation, adaptation and speciation. Heredity 83: 363–372.
- Rockman, M. V., 2012 The QTN program and alleles that matter for evolution: all that's gold does not glitter. Evolution 66: 1–17.
- Rockman, M. V., and L. Kruglyak, 2008 Breeding designs for recombinant inbred advanced intercross lines. Genetics 179: 1069–1078.
- Rockman, M. V., and L. Kruglyak, 2009 Recombinational landscape and population genomics of *Caenorhabditis elegans*. PLoS Genet. 5: e1000419.
- Rockman, M. V., S. S. Skrovanek, and L. Kruglyak, 2010 Selection at linked sites shapes heritable phenotypic variation in *C. ele*gans. Science 330: 372–376.
- Ryu, W. S., and A. D. T. Samuel, 2002 Thermotaxis in *Caenorhabditis elegans* analyzed by measuring responses to defined thermal stimuli. J. Neurosci. 22: 5727–5733.
- Segrè, D., A. DeLuna, G. M. Church, and R. Kishony, 2004 Modular epistasis in yeast metabolism. Nat. Genet. 37: 77–83.
- Seidel, H. S., M. V. Rockman, and L. Kruglyak, 2008 Widespread genetic incompatibility in *C. elegans* maintained by balancing selection. Science 319: 589–594.
- Seidel, H. S., M. Ailion, J. Li, A. van Oudenaarden, M. V. Rockman et al., 2011 A novel sperm-delivered toxin causes late-stage embryo lethality and transmission ratio distortion in *C. elegans*. PLoS Biol. 9: e1001115.
- Stiernagel, T., 1999 Maintenance of *C. elegans*, pp. 51–67 in *C. elegans: A Practical Approach*, edited by I. A. Hope. Oxford University Press, Oxford.
- Weber, K. P., S. De, I. Kozarewa, D. J. Turner, M. M. Babu *et al.*, 2010 Whole genome sequencing highlights genetic changes associated with laboratory domestication of *C. elegans*. PLoS ONE 5: e13922.
- Whitlock, M. C., P. C. Phillips, F. B.-G. Moore, and S. J. Tonsor, 1995 Multiple fitness peaks and epistasis. Annu. Rev. Ecol. Syst. 26: 601–629.
- Wicks, S. R., R. T. Yeh, W. R. Gish, R. H. Waterston, and R. H. Plasterk, 2001 Rapid gene mapping in *Caenorhabditis elegans* using a high density polymorphism map. Nat. Genet. 28: 160–164.
- Wright, S., 1932 The roles of mutation, inbreeding, crossbreeding and selection in evolution. Proc. Sixth Int. Cong. Genet. 1: 356–366.
- Zariwala, H. A., A. C. Miller, S. Faumont, and S. R. Lockery, 2003 Step response analysis of thermotaxis in *Caenorhabditis elegans*. J. Neurosci. 23: 4369–4377.
- Zuk, O., E. Hechter, S. R. Sunyaev, and E. S. Lander, 2012 The mystery of missing heritability: genetic interactions create phantom heritability. Proc. Natl. Acad. Sci. USA 109: 1193–1198.

Communicating editor: D. W. Threadgill