Polygenicity and Epistasis Underlie Fitness-Proximal Traits in the *Caenorhabditis elegans* Multiparental Experimental Evolution (CeMEE) Panel

Luke M. Noble,*[,1] Ivo Chelo,[†] Thiago Guzella,[‡] Bruno Afonso,[†,‡] David D. Riccardi,* Patrick Ammerman,*
Adel Dayarian,[§] Sara Carvalho,[†] Anna Crist,[‡] Ania Pino-Querido,[†] Boris Shraiman,[§,**]
Matthew V. Rockman,*[,1] and Henrique Teotónio[‡,1]

*Center for Genomics and Systems Biology, Department of Biology, New York University, New York 10003, [†]Instituto Gulbenkian de Ciência, P-2781-901 Oeiras, Portugal, [‡]Institut de Biologie, École Normale Supérieure, Centre National de la Recherche Scientifique (CNRS) UMR 8197, Institut National de la Santé et de la Recherche Médicale (INSERM) U1024, F-75005 Paris, France, [§]Kavli Institute for Theoretical Physics and **Department of Physics, University of California, Santa Barbara, California 93106

ORCID IDs: 0000-0002-5161-4059 (L.M.N.); 0000-0002-5414-6625 (I.C.); 0000-0003-0886-8990 (B.S.); 0000-0001-6492-8906 (M.V.R.)

ABSTRACT Understanding the genetic basis of complex traits remains a major challenge in biology. Polygenicity, phenotypic plasticity, and epistasis contribute to phenotypic variance in ways that are rarely clear. This uncertainty can be problematic for estimating heritability, for predicting individual phenotypes from genomic data, and for parameterizing models of phenotypic evolution. Here, we report an advanced recombinant inbred line (RIL) quantitative trait locus mapping panel for the hermaphroditic nematode *Caenorhabditis elegans*, the *C. elegans* multiparental experimental evolution (CeMEE) panel. The CeMEE panel, comprising 507 RILs at present, was created by hybridization of 16 wild isolates, experimental evolution for 140–190 generations, and inbreeding by selfing for 13–16 generations. The panel contains 22% of single-nucleotide polymorphisms known to segregate in natural populations, and complements existing *C. elegans* mapping resources by providing fine resolution and high nucleotide diversity across > 95% of the genome. We apply it to study the genetic basis of two fitness components, fertility and hermaphrodite body size at time of reproduction, with high broad-sense heritability in the CeMEE. While simulations show that we should detect common alleles with additive effects as small as 5%, at gene-level resolution, the genetic architectures of these traits do not feature such alleles. We instead find that a significant fraction of trait variance, approaching 40% for fertility, can be explained by sign epistasis with main effects below the detection limit. In congruence, phenotype prediction from genomic similarity, while generally poor ($r^2 < 10\%$), requires modeling epistasis for optimal accuracy, with most variance attributed to the rapidly evolving chromosome arms.

KEYWORDS genetic architecture; polygenicity; epistasis; experimental evolution; body size; fertility; selfing; GWAS; heritability; quantitative trait; complex trait; QTL; MPP; multiparental populations; Multiparent Advanced Generation Inter-Cross (MAGIC)

MOST measurable features of organisms vary among individuals. Outlining the genetic dimension of this variation, and how this varies across populations and traits, has important implications for the application of genomic data

to predict disease risk and agricultural production, for estimation of heritability, and for understanding evolution (Lynch and Walsh 1998; Barton and Keightley 2002). Complex traits are defined by being multifactorial. They tend to be influenced by many genes and to be plastic in the presence of environmental variation, and the manner in which phenotypic variation emerges from the combined effects of causal alleles is rarely clear. Although phenotype prediction and some aspects of evolution can often be well approximated by considering additive effects alone, nonadditive interactions between alleles at different loci (with marginal additive effects) may explain a large fraction of trait variation yet remain undetected due to low statistical power (Phillips 2008). Adding further complication, one cannot usually assume that

genetic and environmental effects are homogeneous or independent of one another (Barton and Turelli 1991; Félix and Barkoulas 2015), nor that the genetic markers used for mapping quantitative trait loci (QTL) are faithfully and uniformly associated with causal alleles (Yang *et al.* 2010; Speed *et al.* 2012).

Human height, for example, is the canonical quantitative trait, an easily measured, stable attribute with high heritability ($\sim$80%) when measured in families (Galton 1886); Fisher 1930; Visscher *et al.* 2010). Hundreds of common QTL [minor allele frequency (MAF) > 5%] of small effect have been detected by genome-wide association studies (GWAS), explaining in sum a small fraction ($\sim$20%) of heritability (Wood *et al.* 2014). A recent study with $> 7 \times 10^5$ people showed that close to 100 uncommon, larger effect QTL (0.1% < MAF < 5%) explain a mere extra 5% of heritability (Marouli *et al.* 2017). It has taken methods of genomic selection in animal breeding, and dense genetic marker information (Meuwissen *et al.* 2001; Meuwissen and Goddard 2010), to show that common, undetected QTL can potentially explain a large fraction of the variability in human height and common diseases (Yang *et al.* 2010; Speed *et al.* 2016). Thus, in perhaps many cases, the so-called problem of the missing heritability may be synonymous with high polygenicity (Hill *et al.* 2008; Manolio *et al.* 2009). The contribution of statistical epistasis to variation in human height (as a variance component orthogonal to additive and dominance effects) is likely to be modest (Visscher *et al.* 2010), but nonetheless important for predicting individual phenotypes and for understanding the long-term response to selection. Molecular genetics and biochemistry suggest that nonadditivity is ubiquitous [an idiosyncratic sample of studies showing effects of natural variation on organismal traits includes Mukai (1967), Whitlock and Bourguet (2000), Bonhoeffer *et al.* (2004), Carlborg *et al.* (2006), Shao *et al.* (2008), Zwarts *et al.* (2011), Gaertner *et al.* (2012), Corbett-Detig *et al.* (2013), Huang *et al.* (2014), Bloom *et al.* (2015), Monnahan and Kelly (2015), Paaby *et al.* (2015), Chirgwin *et al.* (2016), and Forsberg *et al.* (2017)], but the general importance of epistasis for explaining missing heritability, shaping fitness landscapes, maintaining genetic diversity, facilitating the evolution of recombination, and generating the additive genetic variance on which selection can act, continues to be debated (Cheverud and Routman 1995; Wolf *et al.* 2000; Hill *et al.* 2008; Phillips 2008; Rockman 2012; Hansen 2013; Hemani *et al.* 2013; Weinreich *et al.* 2013; Mackay *et al.* 2014; Barton 2017).

Alongside GWAS, inbred line crosses in model systems have been instrumental for our understanding of the genetics of complex traits, given the opportunity for control of confounding environmental covariates and accurate measurement of breeding values. Crosses among multiple parental strains in particular—such as those now available for mice (Churchill *et al.* 2004), *Drosophila* (Macdonald and Long 2007), maize (Buckler *et al.* 2009; McMullen *et al.* 2009), wheat (Huang *et al.* 2012; Mackay *et al.* 2014; Thepot *et al.*

2015), rice (Bandillo *et al.* 2013), tomato (Pascual *et al.* 2015), and *Arabidopsis* (Kover *et al.* 2009), among others—have been developed to better sample natural genetic variation. Greater variation also allows the effects of multiallelic loci to be studied and, subject to effective recombination, improved QTL resolution. If large populations and random mating are imposed for long periods, gains in resolution can be dramatic (Valdar *et al.* 2006; Rockman and Kruglyak 2008), though this comes with increased opportunity for selection to purge diversity (*e.g.*, Rockman and Kruglyak 2009; Baldwin-Brown *et al.* 2014).

Better known as a model for functional biology (Corsi *et al.* 2015), the nematode *Caenorhabditis elegans* has also contributed to our understanding of complex traits and their evolution. *C. elegans* shows extensive variation in complex traits, and sex-determination and breeding mode (selfing and outcrossing) can be genetically manipulated at will (Johnson and Wood 1982; Hodgkin and Doniach 1997; Gems and Riddle 2000; Knight *et al.* 2001; Barrière and Félix 2005; Gutteling *et al.* 2007; Diaz and Viney 2014; Gray and Cutter 2014; Teotónio *et al.* 2017). QTL for traits such as embryonic lethality (Rockman and Kruglyak 2009), pesticide resistance (Ghosh *et al.* 2012), and telomere length (Cook *et al.* 2016) have been found by association studies in an expanding panel of inbred wild isolates, the *C. elegans* natural diversity resource [CeNDR; https://elegansvariation.org/ and Cook *et al.* (2017)]. Complex traits have also been studied using collections of recombinant inbred (Rockman and Kruglyak 2009) and introgression lines (RILs and ILs; Doroszuk *et al.* 2009) derived by crossing the laboratory domesticated N2 strain (Sterken *et al.* 2015) and the divergent Hawaiian wild isolate CB4856 (*e.g.*, Andersen *et al.* 2014, 2015), or other biparental crosses (*e.g.*, Duveau and Félix 2012; Noble *et al.* 2015). GWAS and line crosses have given insights into how natural selection has shaped phenotypic variation in *C. elegans* and related nematodes. For example, an N2/CB4856 RIL panel has been used to argue that selection on linked sites largely explains the distribution of QTL effects for mRNA abundance (Rockman *et al.* 2010). Lastly, *C. elegans* is also one of the main models for experimental evolution (Gray and Cutter 2014; Teotónio *et al.* 2017). Mutation accumulation line panels in particular have long been used to estimate mutational heritability (Estes and Lynch 2003; Baer *et al.* 2005; Estes 2005; Baer 2008; Halligan and Keightley 2009; Phillips *et al.* 2009) and to argue that standing levels of genetic variation in natural populations for complex traits can be explained by a mutation–selection balance (Etienne *et al.* 2015; Farhadifar *et al.* 2016). As yet, existing *C. elegans* RIL panels are limited to biparental crosses, and yield coarse QTL mapping resolution.

A prominent characteristic of *C. elegans* is its mixed androdioecious reproductive system, with hermaphrodites capable of either selfing, from a cache of sperm produced late in larval development (Hirsh *et al.* 1976), or outcrossing with males (Maupas 1900). Sex determination

is chromosomal, with hermaphrodites XX, and XO males maintained through crosses and rare X chromosome non-disjunction during hermaphrodite gametogenesis (Nigon 1949). Because males are typically absent from selfed broods but are half the progeny of a cross, twice the male frequency in a population is the expected outcrossing rate (Stewart and Phillips 2002; Cutter 2004). Natural populations have low genetic diversity and very high linkage disequilibrium (LD), with generally weak global population structure and high local diversity among typically homozygous individuals at the patch scale (Barrière and Félix 2005, 2007; Cutter *et al.* 2009). Average single-nucleotide polymorphism (SNP) diversity is in the order of 0.3% (Cutter 2006) though highly variable across the genome, reaching $\geq 16\%$ in some hypervariable regions (Thompson *et al.* 2015). Low diversity and high LD is due to the predominance of inbreeding by selfing, which reduces the effective recombination rate and elevates susceptibility to linked selection (Graustein *et al.* 2002; Rockman *et al.* 2010; Andersen *et al.* 2012). Crosses between wild isolates have revealed outbreeding depression (Dolgin *et al.* 2007; Chelo *et al.* 2014), presumably due to disruption of epistatic allelic interactions (*e.g.*, Gaertner *et al.* 2012).

Perhaps as a consequence of low but significant outcrossing (and a metapopulation demographic structure), several loci have been found to be under some form of balancing selection (*e.g.*, Seidel *et al.* 2008; Ghosh *et al.* 2012; Greene *et al.* 2016). Moreover, evolution experiments involving crosses among multiple strains have shown that high rates of outcrossing can persist given heritable variation for mating traits (Anderson *et al.* 2010; Teotónio *et al.* 2012; Masri *et al.* 2013), and that frequent outcrossing in moderately-sized populations may have facilitated loss of genetic diversity by partial selective sweeps, with excess heterozygosity maintained by epistatic selection on overdominant loci (*e.g.*, Chelo and Teotónio 2013; Chelo *et al.* 2014).

This molecular and population genetic foundation suggests that study of *C. elegans* may be fruitful for our understanding of the contribution of within- and between-locus nonadditive interactions to complex traits and their evolution. Here, we present a panel of 507 genome-sequenced RILs obtained by intercrossing 16 wild isolates, and culturing at high outcrossing rates in populations of $\approx 10^4$ for 140–190 generations of experimental evolution, followed by inbreeding by selfing for 13–16 generations. The *C. elegans* Multiparental Experimental Evolution (CeMEE) RIL panel complements existing *C. elegans* mapping resources by providing fine-mapping resolution and high nucleotide diversity. Using simulations, we show that the CeMEE panel can give gene-level resolution for common QTL with effects as low as 5%. Using subsets of RILs, we investigate the genetic basis of two fitness components, fertility and adult hermaphrodite body size, by variance decomposition under additive and additive-by-additive epistatic models, and by genome-wide one- and two-dimensional (1D and 2D, respectively) association testing. We find that the genetic basis of

both traits, particularly fertility, is highly polygenic, with a significant role for interactions of large effect.

Materials and Methods

CeMEE derivation

The panel was derived in three stages (Figure 1). First, 16 wild isolates (AB1, CB4507, CB4858, CB4855, CB4852, CB4856, MY1, MY16, JU319, JU345, JU400, N2 (ancestral), PB306, PX174, PX179, and RC301; obtained from the *Caenorhabditis* Genetics Center) were inbred by selfing for 10 generations to ensure homozygosity, then intercrossed to funnel variation into a single multiparental hybrid population, as described in Teotónio *et al.* (2012). Each of the four funnel phases comprised multiple pairwise, reciprocal crosses at moderate population sizes (see Supplemental Material, and figure S1 and supporting information of Teotonio *et al.* (2012) for full details of replication and population sizes).

Second, the multiparental hybrid population was evolved for 140 discrete generations at population sizes of $N \approx 10^4$ (outcrossing rate $\approx 0.5, N_e \approx 10^3$), to obtain the A140 population [as reported in Teotónio *et al.* (2012), Chelo and Teotónio (2013), and Chelo *et al.* (2013)]. Sex-determination mutations were then mass introgressed into the A140, while maintaining genetic diversity, to generate monoecious (obligately selfing hermaphrodites) and trioecious (partial selfing with males, females, and hermaphrodites) populations, as detailed in Theologidis *et al.* (2014). Further replicated experimental evolution was carried out for 50 generations under two environmental regimes: (1) a control regime (conditions as before), with the wild-type androdioecious reproductive system (CA50 collectively); and (2) a gradual exposure to an increasing gradient of NaCl, from 25 mM (standard NGM-lite medium; United States Biological) to 305 mM until generation 35 and thereafter, varying reproductive system (GX50, where X is androdioecious, monoecious, or trioecious). Although trioecious populations started evolution with only 0.1% of hermaphrodites, by generation 50 they were abundant [50%; see figure S7 in Theologidis *et al.* (2014)]. Androdioecious populations maintained outcrossing rates of > 0.4 until generation 35, soon after losing males to finish with an outcrossing rate of ~ 0.2 by generation 50 [figure S5 in Theologidis *et al.* (2014)]. This complex experimental evolution scheme was designed to study the effects of reproductive system on the genetics and evolution of complex traits; however, here we consider this structure only in so far as it is relevant to the mapping of quantitative traits in the panel as a whole.

Finally, hermaphrodites were inbred by selfing to obtain RILs. Population samples ($> 10^3$ individuals) were thawed from $-80°$ and maintained under standard laboratory conditions for two generations. At the third generation, single hermaphrodites were picked at the late-third to early-fourth larval stage (L3/L4) and placed in wells of 12-well culture plates, containing M9 medium (25 mM NaCl) seeded with
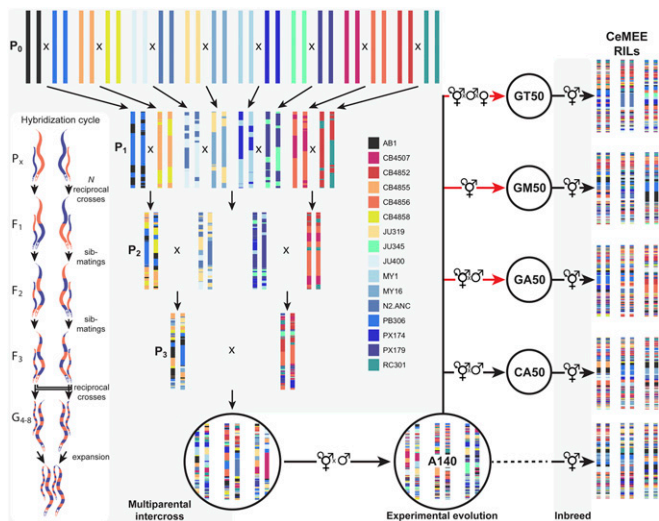
Figure 1 CeMEE derivation. The multiparental intercross funnel phase comprised four stages of pairwise crosses and progeny mixing, carried out in parallel at controlled population sizes. One hybridization cycle for a single-founder cross is inset at left: in each cycle, multiple reciprocal crosses were initiated, increasing in replicate number and census size each filial generation. F_1 and F_2 progeny were first sib-mated, then reciprocal lines were merged by intercrossing the F_3 and expanding the pooled G_4 (for three to four generations) before commencing the next reduction cycle. The resulting multiparental hybrid population was archived by freezing, and samples were thawed and maintained for 140 nonoverlapping generations of mixed selfing and outcrossing under standard laboratory conditions to generate the A140 population. Hermaphrodites were then sampled from the A140 and selfed to generated the A140 RILs. Additionally, the outbred A140 population was evolved for a further 50 generations under the same conditions (CA) or under adaptation to a salt gradient with varying sex ratios (GT, GM, and GA lines; Theologidis *et al.* 2014). See *Materials and Methods* for description of subpanels, and Teotónio *et al.* (2012) for details of replicate numbers and population sizes for each funnel generation. CA, control adapted lines; CeMEE, *C. elegans* multiparental experimental evolution panel; GA, gradual adaptation androdioecious; GM, gradual adaptation monoecious; GT, gradual adaptation trioecious; RILs, recombinant inbred lines.

Escherichia coli. Lines were propagated at 20° and 80% relative humidity (RH) by transferring a single L3/L4 individual for 16 (A140 population) or 13 generations (4–7 days between transfers). At each passage, parental plates were kept at 4° to prevent growth until offspring production was verified, and in cases of failure two additional transfers were attempted before declaring line extinction. Inbreeding was done in several blocks from 2012 to 2016, in two different locations. A total of 709 RILs were obtained and archived at −80°. Full designation of CeMEE RILs and subpanels are in File S1 and File S2.

Sequencing and genotyping

Full details of sequencing, genotype calling, and variant filtration can be found in the supplemental material. In brief, founders were sequenced to $\geq 30\times$ depth with Illumina 50 or 100 bp paired-end reads, and variants were called against the WS245 *C. elegans* N2 reference genome (GATK 3.3-0 HaplotypeCaller; McKenna *et al.* 2010). After depth, quality, zygosity, and frequency filtering, we arrived at a final set of 388,201 founder SNP markers at which to genotype RILs.

RILs were sequenced with 100 or 150 bp paired-end reads to a mean depth of $5.1\times$. Genotypes were imputed by Hidden Markov Model (HMM), considering the 16 founder states and mean base qualities of reads. After removing closely related lines, we retained 178 A140 RILs, 118 CA50 RILs (from three replicate populations), 127 GA50 RILs (three replicates), and 79 GT50 RILs (two replicates). The 98 GM50 RILs (two replicates) are derived from monoecious populations and are highly related on average, grouping together into a small number of "isotypes." To prevent the introduction of strong structure, we discarded all but five below a panel-wide pairwise identity threshold for the purposes of trait mapping (taking the line with greatest sequence coverage for each isotype, grouped by mean pairwise identity among lines of all other subpanels + 5 SD). In total, the CeMEE comprises 507 RILs from five subpanels, with 352,583 of the founder markers segregating within it. Raw and filtered founder variant calls are in File S3 and File S4, and imputed RIL genotypes are in File S5.

We estimated residual heterozygosity for 25 A140 lines sequenced to $> 20\times$ coverage (single sample calls using GATK 3.3-0 HaplotypeCaller, variant filtration settings $MQ < 50.0$ || $DP < 5$ || $MQRankSum < -12.5$ || $SOR < 6$ || $FS > 60.0$ || $ReadPosRankSum < -8.0$ || $QD < 10.0$ || $DP > mean \times 3$). Mean heterozygosity in these lines at founder sites is 0.095% (SD 0.042%, range 0.033–0.18%).

Genetic marker sets

Four subsets of the 352,583 founder SNPs segregating in the CeMEE panel are used for analysis (referenced in the corresponding sections), which we define here:

1. Subset of 248,668 markers used for GWAS with MAF > 0.05 in phenotyped lines.
2. Subset of 88,508 markers pruned of strong local LD (generated by LDAK in a two-pass window-based filtering on $r^2 < 0.98$ (see *Heritability and phenotype prediction*), used for analysis of interchromosomal LD and panel structure.
3. Subset of 4960 markers used for 2D testing, with MAF > 0.05, weak local LD (Plink –indep-pairwise, window = 200 kb, step = 10, $r^2 < 0.5$), ≤ 5 missing or ambiguous imputed genotypes, and filtering across marker pairs for the presence of all four two-locus homozygote classes at a frequency of ≥ 3 in at least one test.
4. Subset of 256,535 diallelic sites shared between the CeMEE and CeNDR panel with no missing or heterozygous calls.

CeMEE genetic structure

Differentiation from natural isolates and founders: We compared similarity within and between the CeMEE RILs and 152 sequenced wild isolates from the CeNDR panel (release

20160408). The distributions for all pairwise genotype and haplotype (% identity at 0.33 cM scale in F_2 map distance) distances are plotted in Figure S1 in File S22, using marker set 4.

LD (r^2) was computed for founders and CeMEE RILs at the same set of sites (marker set 2, additionally filtered to MAF > 1/16, then subsampled by a factor of 10 for computational tractability), and plotted against genetic distances [obtained by linear interpolation from the N2/CB4856 map, scaled to F_2 distances (Rockman and Kruglyak 2009)]. To assess the presence of subtle, long-range LD in the form of interchromosomal structure, we compared mean r^2 among chromosomes to a null distribution generated by permutation, where associations between chromosomes are randomized within each RIL genome and the contribution of allele frequency differentiation between subpanels is controlled. In each permutation ($n = 5000$), RIL genotypes (marker set 2) were randomly subsampled to equal size across chromosomes, split by chromosome, then shuffled within each subpanel, before taking the mean correlation across chromosomes as the test statistic (or omitting all single and pairwise chromosome combinations). The effect of local LD pruning is to reduce the weighting of long haplotypes in strong LD, to better assay weak interactions involving loci distributed throughout the majority of the genome. Permutation code is in File S6 (interchromLD.R).

Reconstruction of ancestral haplotypes and genetic map expansion:

For each RIL, founder haplotypes were inferred with the RABBIT HMM framework implemented in Mathematica (Zheng *et al.* 2015), conditioning on the recombination frequencies observed for the N2/CB4856 RILs (scaled to F_2 map length; coordinates are in File S7 WS220_geneticMap.txt) (Rockman and Kruglyak 2009). Realized map expansion was estimated by maximum likelihood for each chromosome, before full marginal reconstruction (explicitly modeling recombination on the X chromosome and autosomes) using posterior decoding under the fully-dependent homolog model (depModel). Under this model, appropriate for fully inbred diploids, chromosome homologs are assumed to have identical ancestral origins (prior identity by descent probability $f = 1$), and the recombination junction density (transition probability) is given by the estimated map expansion (Ra) and genotyping error rates (set to 5×10^{-5} for founders and 5×10^{-3} for RILs based on likelihood from a parameter sweep). Sites called as heterozygous or missing in the founders, or unresolved to $[0, 1]$ by the genotype imputation HMM in RILs, were set to NA (missing data) before reconstruction. To summarize performance, per marker posterior probabilities were filtered to > 0.2, and haplotype lengths and breakpoints were estimated from run lengths of marker assignments, taking the single best haplotype (if present), maintaining haplotype identity (if multiple assignments of equal probability), or the first among equals otherwise.

To test reconstruction accuracy as a function of haplotype length, we performed matched simulations varying only the number of generations of random mating (code in File S8 RABBIT_simulations.R). Starting from a single population representing all founders [$N = 1000$, corresponding to the expected N_e during experimental evolution (Chelo and Teotónio 2013)], mating occurred at random with equal contribution to the next generation. Recombination between homologous chromosomes occurred at a rate of 50 cM, with full crossover interference, and the probability of meiotic crossover based on distances between marker pairs obtained by linear interpolation of genetic positions (Rockman and Kruglyak 2009). For each chromosome, 10 simulations were run sampling at 10, 25, 50, 100, 150, 200, 250, and 300 generations, and haplotype reconstruction was carried out as above. Maximum likelihood estimates of realized map expansion for simulations were used to calibrate a model for prediction of the effective number of generations in the RILs. With increasing generation number, Ra was progressively underestimated due to unresolved small recombination events (*e.g.*, 14% mean deviation at generation 300). Given this, we used a second-degree polynomial regression of Ra on the known number of generations, which was significantly preferred over a linear fit by likelihood ratio test (LRT).

Population stratification:

Population stratification was assessed using principal component (PC) decomposition, and supervised and unsupervised discriminant analysis of PCs (DAPC; Jombart *et al.* 2010). In all cases, decomposition was of genotypes pruned of strong local LD (marker set 2), mean centered, and scaled to unit variance.

Of the first 50 PCs, 10 are individually significantly associated with subpanel identity by ANOVA (linear regression of each PC on subpanel identity, tested against an intercept only model by LRT, $P < 0.05$ after Bonferroni correction). Seven of the top 10 PCs are significant, though others up to 38 are also associated, showing that multiple sources of structure contribute to the major axes of variation.

For DAPC [R package adegenet, Jombart (2008)], we used 100 rounds of 10-fold cross-validation to determine the number of PCs required for optimal subpanel assignment accuracy (the mean of per-group correct assignments). This value (40 PCs) was then used to infer groups by unsupervised k-means clustering (default settings of 10 starts, 10^5 iterations) with the number of groups selected on the Bayesian Information Criterion (BIC).

Phenotyping

Fertility:

In the experimental evolution scheme under which the CeMEE RILs were generated, a hermaphrodite's contribution to the next generation is the number of viable embryos that survive bleaching (laid, but unhatched, or *in utero*) that subsequently hatch to L1 larvae 24-hr later. We treat this phenotype as fertility, and measured it for individual worms of 230 RILs. Full details are provided in File S22. In brief, we

used manually-scored plate-based assays of the number of viable embryos produced by single adult hermaphrodites, with two independent plates for most RILs, which we consider as replicates for estimation of repeatability (see below). In total, the median number of measurements per line was 43 (range 4–84). Final trait values were the Box–Cox transformed line coefficients from a Poisson generalized linear model (log link) with fixed categorical effects of plate row, column, and edge (exterior rows and columns), and the count of offspring per worm as response variable (model S1 in supplemental material). Data for 227 RILs passed filtering (raw data are in File S9, model coefficients are in File S10), coming from RILs of three subpanels (170 A140, 45 GA50, and 12 GT50). Subpanel explains 4% of the variance in this trait, with GA50 RILs having higher mean fertility than the A140 (linear regression of trait values on subpanel identity, regression coefficient = 0.43, $P = 0.01$, see Figure S2 in File S22).

Adult hermaphrodite body size: The area of adult hermaphrodite worms was measured using a Multi-Worm Tracker (Swierczek *et al.* 2011). Data were generated in two laboratory locations over several years, recording the relative humidity and temperature at the time of assay (see supplemental material for full details). Final trait values were the Box–Cox transformed line coefficients from a linear model incorporating fixed effects of year, nested within location, and humidity and temperature, nested within location (model S3 in Supplemental Methods). Data for 410 RILs passed filtering, with two independent thaw blocks for most RILs (raw data are in File S11 and final trait values are in File S12). Data come from RILs of three subpanels (165 A140, 118 CA50, and 127 GA50), which explain 17% of variance in this trait. GA50 RILs are much larger than the A140 (regression coefficient = 0.94, $P < 10^{-16}$; see Figure S2 in File S22), and this is not driven by technical covariates: data acquisition for A140 RILs and GA50 RILs was relatively balanced with respect to location and time, and GA50 RILs are significantly larger across all five laboratory/year blocks.

Fertility and body size show significant phenotypic and genetic correlations [Figure S2 in File S22; see also Poullet *et al.* (2016)], justifying the latter being considered a fitness-proximal trait. For 202 lines with data for both traits, the phenotypic correlation = 0.35 (Spearman's ρ for the final trait model coefficients used for QTL mapping, $P = 2.3 \times 10^{-7}$), and genetic correlation $r_G = 0.54$ ($\sigma_{s,f}/\sqrt{\sigma_s^2 \sigma_f^2}$, where $\sigma_{s,f}$ is genetic covariance between size and fertility, was estimated by restricted/residual maximum likelihood (REML) [R package sommer, Covarrubias-Pazaran (2016)] using unweighted additive genetic similarity **A** (see below).

Heritability and phenotype prediction

Repeatability: Repeatability was estimated from ANOVA of the line replicate means for each trait as $R = \sigma_a^2/(\sigma_a^2 + \sigma_e^2)$, where $\sigma_a^2 =$ (mean square among lines − mean square error)/n_0, and n_0 is a coefficient correcting for a varying

number of observations (1−4 plate means) per line (Lessells and Boag 1987; Sokal and Rohlf 1995). Assuming equal variance and equal proportions of environmental and genetic variance among replicates, R represents an upper bound on broad-sense heritability (Falconer 1981; Hayes and Jenkins 1997). Fertility data were square root-transformed to decouple the mean and variance.

Assumptions: In inbred, isogenic lines, broad-sense heritability can also be estimated by linear mixed-effects model (LMM) from the covariance between genetic and phenotypic variances. However, the measurement of genetic similarity is subject to a number of assumptions and is (almost) always, at best, an approximation (Speed and Balding 2015).

A first assumption is that all markers are the causal alleles of phenotypic variation. However, it is unavoidable that markers tag the (unknown) causal alleles to different degrees due to variable LD. A second, usually implicit, assumption in calculating genetic similarity is the weight given to markers as a function of allele frequency. Greater weight has typically been given to rare alleles in human research, which has support under scenarios of both selection and neutrality (Pritchard 2002). A third assumption, related to the first two, is the relationship between LD and causal variation. If the relationship is positive—causal variants being enriched in regions of high LD—then heritability estimated from all markers will be upwardly biased, since the signal from causal variation contributes disproportionately to genetic similarity (Speed *et al.* 2012).

The use of whole-genome sequencing largely addresses the first assumption, given (as here) very high marker density and an accurate reference genome, although in the absence of full *de novo* genomes from long-read data for each individual, the contribution of large-scale copy number and structural variation, and new mutation, will remain unknown. To account for the second and third assumptions, we used LDAK (v5.0) to explicitly account for LD in the CeMEE (decay half-life = 200 kb, min-cor (minimum squared correlation coefficient) = 0.005, min-obs (minimum percentage of non-missing data per marker) = 0.95; Speed *et al.* 2012). Heritability estimates were not sensitive to variation in the decay parameter over a 10-fold range or to the measurement unit (physical or genetic), and we used physical distance. Across the set of 507 RILs, 88,508 segregating markers were used after local LD-based pruning (marker set 2) and, of these, 22,984 markers received nonzero weights (File S13). LD weighting can magnify the effects of genotyping errors. We tested the effect of excluding 17,740 markers with particularly low local LD (mean r^2 over a 20-marker window < 0.3, or the ratio of mean r^2 to that of the window mean < 0.3) before estimation of LD weights. Heritability estimates were largely unchanged (within the reported intervals), as were our general conclusions on variance components and model performance.

Modeling: Given m SNPs, genetic similarity is calculated by first scaling S, the $n \times m$ matrix of mean centered genotypes,

where $S_{i,j}$ is the number of minor alleles carried by line i (of n) at marker j and frequency f, to give X:

$$X_{i,j} = \left(S_{i,j} - 2f_j\right) \times \left(2f_j(1-f_j)\right)^{\frac{\alpha}{2}}; \qquad (1)$$

The additive genetic similarity matrix (GSM) **A** is then $\mathbf{XX}^T/m$. Here, α scales the relationship between allele frequency and effect size (Speed *et al.* 2012), $\alpha = -1$ corresponds to the assumption of equal variance explained per marker (an inverse relationship of effect size and allele frequency), while common alleles are given greater weight at $\alpha > 0$. We tested $\alpha \in [-1.5, -1, -0.5, 0, 0.5, 1]$ and report results that maximized prediction accuracy. With Y the mean centered vector of n phenotype values scaled to unit variance, the additive model fit for estimating genomic heritability (h^2) is then:

$$Y = \sum_{}^{m} A\beta + e, \\ \text{with } \beta \sim \mathcal{N}\left(0, \sigma_g^2\right), \ e \sim \mathcal{N}\left(0, \sigma_e^2\right) \qquad (2)$$

where β represents random SNP effects capturing genetic variance σ_g^2, and e is the residual error capturing environmental variance σ_e^2. Given Y and **A**, heritability can be estimated from REML estimates of genetic and residual variance as $h^2 = \sigma_g^2/(\sigma_g^2 + \sigma_e^2)$. Note that we use the terms h^2 and genomic heritability interchangeably here for convenience, although in some cases nonadditive covariances are included. We assume RILs are fully inbred.

The existence of near-discrete recombination rate domains across chromosomes has led to a characteristic structure of nucleotide variation, correlated with gene density and function (Cutter *et al.* 2009). Variation also varies widely among chromosomes (Rockman *et al.* 2010; Andersen *et al.* 2012). This heterogeneity is not captured by aggregate genome-wide similarity with equal marker weighting (Speed *et al.* 2012; Goddard *et al.* 2016). To better reflect observed LD, markers were first weighted by the amount of genetic variation they tag along chromosomes (Speed *et al.* 2012). Given m weights, $w_i, \ldots, w_m$, genetic effects for the basic model become:

$$\beta \sim \mathcal{N}\left(0, w\sigma_g^2/W\right) \qquad (3)$$

where W is a normalizing constant. Second, we jointly measured the variance explained by individual chromosomes (and by genetic variation in recombination rate domains within each chromosome), which can potentially improve the precision of heritability estimation if causal variants are not uniformly distributed by allowing variance to vary among partitions. Third, we tested epistatic as well as additive genetic similarity with (1) the entrywise (Hadamard) product of additive GSMs, giving the probability of allele pair sharing (Henderson 1985; Jiang and Reif 2015), (2) higher exponents up to fourth-order interactions, and (3) haplotype-based similarity at multi-gene scale. Additional

similarity components (additive or otherwise) are added as random effects to the above model to obtain independent estimation of variance components (see supplemental materials for details).

Model fit was assessed by phenotype predictions from leave-one-out cross-validation, calculating the genomic best linear unbiased prediction (GBLUP) (Meuwissen *et al.* 2001; VanRaden 2008) for each RIL and returning the squared correlation coefficient (r^2) between observed and predicted trait values (carried out in LDAK). To avoid bias associated with sample size, all models were unconstrained (nonerror variance components were allowed to vary outside 0–1 during convergence) unless otherwise noted, which generally gave better likelihood for multi-component models.

GWAS

1D tests: For single trait, single marker association, we fitted LMMs:

$$Y = \beta X + g + e, \ \text{with } g \sim \mathcal{N}\left(0, \sigma_g^2 A\right), \ e \sim \mathcal{N}\left(0, \sigma_e^2\right) \qquad (4)$$

where X is the matrix of fixed effects (SNP genotype) and β is the effect on phenotypic variation that is estimated. g are the random effects describing genetic covariances (Equation 3) accounting for nonindependence among tests due to an assumed polygenic contribution to phenotype, with **A** the $n \times n$ GSM from the most predictive additive fit found for each trait, and e is residual error. The above model was compared to a null model excluding genotype effects by LRT (fit using the LIMIX Python package, https://github.com/limix/limix). GWAS *P*-values for size and fertility are in File S14, using genetic similarities in File S15 and File S16.

To assess the mapping resolution and power of the CeMEE panel, we carried out GWAS according to the model above for simulated phenotypes. We simulated a single additive locus (h^2 from 1 to 30%) and a background polygenic component of equal variance (scenarios of 10, 100, or 1000 loci), chosen at random from SNPs with MAF > 0.05, with genetic and environmental effect sizes drawn independently from the standard normal distribution (code is in File S17, GWAS_simulations). GWAS was carried out 1000 times for each scenario, controlling for relatedness with LD-weighted additive genetic similarity ($\alpha = -0.5$). Power was estimated from a binomial generalized linear model considering all three polygenic scenarios together. Precision, the fraction of significant QTL that are true positives, was assessed after masking a 1-cM window around the simulated causal SNP. Detection intervals around QTL were defined as a drop in the logarithm of odds (LOD) score of 2 (Morton 1955), and were calculated from similarly powered markers with $\geq$ MAF, with *P*-values converted to LOD scores as $\chi^2/2 \times log(2)/log_{10}(2)$ (Nyholt 2000). True positives were defined as cases where the exact simulated site was detected, and false positives were 2-LOD drop QTL among all other markers detected at a 5% significance permutation threshold (see below).

All 507 lines were used for simulation, and all GWAS tested 248,668 markers with MAF > 0.05 (marker set 1; code is in File S18, GWAS_traits.py). Significance thresholds were established by permutation (Anderson and Ter Braak 2003), with phenotypes generated by permuting phenotype residuals given the estimated relatedness among lines to ensure exchangeability in the presence of polygenic causal effects or structure ($\mathbf{A}$), using the R package mvnpermute (Abney 2015). Significance level α is the corresponding percentile of the minimum P-values from 1000 permutations.

Given correlation between traits, we also tested phenotype residuals for each trait after linear regression on the other, and a multi-trait LMM fitting general and specific effects. No markers were significant in any case (analysis not shown).

2D tests: We tested for epistasis over a reduced search space (marker set 3), on the assumption of complete homozygosity, for a total of 19,913,422 marker pairs (inter- and intrachromosomal). We used a two-level hierarchical procedure, first testing a full linear model ($M1$, main and interaction effects) against a reduced model ($M0$, intercept only) by ANOVA, taking as summary statistic the P-value from an LRT. Significance at level 1 was tested against a null distribution generated by full phenotype permutation (*i.e.*, no additive or interaction effects), with $\alpha = 10\%$ from the minimum values seen for each chromosome pair ($n > 5000$ permutations). We then tested the interaction term specifically for marker pairs significant at level 1 using a parametric bootstrap (Bůžková *et al.* 2011): $M1$ was fitted to responses sampled with replacement from $M2$ ($n = 10,000$), taking the interaction P-value as test statistic, and then comparing the observed statistic to the null distribution (significance declared at $P < 0.01$). Code is in File S19 2D_hierarchical.py.

$$
\begin{aligned}
M0: & \quad Y = \mu + e \\
M1: & \quad Y = \mu + \beta_i X_i + \beta_j X_j + \beta_{ij} X_{ij} + e \\
H0_{(1)}: & \quad \beta_i = \beta_j = \beta_{ij} = 0 \\
M2: & \quad Y = \mu + \beta_i X_i + \beta_j X_j + e \\
H0_{(2)}: & \quad \beta_{ij} = 0
\end{aligned}
\tag{5}
$$

We initially ignored relatedness for 2D testing, then fitted LMMs as above (Equation 2) with genetic covariance $\mathbf{A}$ for candidate interactions (R package hglm; Shen *et al.* 2014). From eight candidate interactions for size, we excluded two for which interaction P-values by LMM were almost an order of magnitude higher. The six remaining candidates changed little (three were lower by LMM). For fertility, interaction P-values were largely insensitive to relatedness (lower for six of eight cases by LMM). Models were also fitted to raw trait values (in addition to the power transformed values) to assess scale effects. One interaction for fertility was significant for transformed values only and was excluded. The amount of phenotypic variance explained by interactions for each trait was estimated by linear model adjusted R^2, jointly fitting all main and two-locus interactions. These estimates were similar to those from LMM variance components, fitting random effects corresponding to additive and additive-by-additive genetic

similarity separately at candidate interactions and background markers (point estimates were 6% lower for size and $< 1\%$ lower for fertility).

We also tested for excess weaker polygenic interactions by taking the sum of log likelihood ratios (LRs) for each marker against all other markers on one other chromosome (2D sum tests). Significance was tested at a single threshold (LR > 16, around the maximum value seen among pairwise interaction null P-values), using the equivalent of the above hierarchical procedure: LRs for $M1$ *vs.* $M0$ were first summed for each marker and compared to a null distribution generated by full phenotype permutation. Candidate markers significant at level 1 ($\alpha = 0.01$) were then tested for significance of the interaction terms against a distribution of LR sums from null additive models for tests with LR > 16 in the observed data. This was repeated 1000 times, with permutation order fixed across bootstraps to maintain correlation structure, and significance was declared at $P < 0.01$. Code is in File S20, 2D_sumLRbootstrap.R.

Data availability

Sequence data are available from the National Center for Biotechnology Information Sequence Read Archive under BioProject PRJNA381203. Raw and processed phenotype and genotype data, and analysis scripts are provided as supplemental material (see DataDocument) and archived in FigShare under DOI: 10.6084/m9.figshare.5466574.v1. RILs are available from the authors.

Results and Discussion

CeMEE differentiation from natural populations

The CeMEE panel of RILs draws variation from 16 founders, and shuffles the diversity they contain through > 150 generations of predominant outcrossing at moderate population size. Isolates used to create the panel together carry $\sim 25\%$ of single-nucleotide variants known to segregate in the global *C. elegans* population (CeNDR; Cook *et al.* 2017). The founders vary considerably in distance to the N2 reference strain, with the Hawaiian CB4856 and German MY16 isolates together contributing over half of all markers, while CB4507 is closely related to N2 (Figure S4A in File S22). Comparison of pairwise genetic distances in the CeMEE and 152 sequenced wild isolates (including a small number of more recently isolated, highly divergent lines) illustrates the scale of novelty generated by the multiparental cross (Figure S1 in File S22). The CeMEE RILs occupy a substantial subspace of CeNDR genotypic diversity, without the extensive haplotype sharing among wild isolates and with the creation of many new multigenic haplotypes (Figure S1B in File S22).

CeMEE differentiation from parental founders

Since *C. elegans* natural isolates suffer from outbreeding depression (Dolgin *et al.* 2007), the mixing phase is expected to generate high variance in fitness which, channeled through
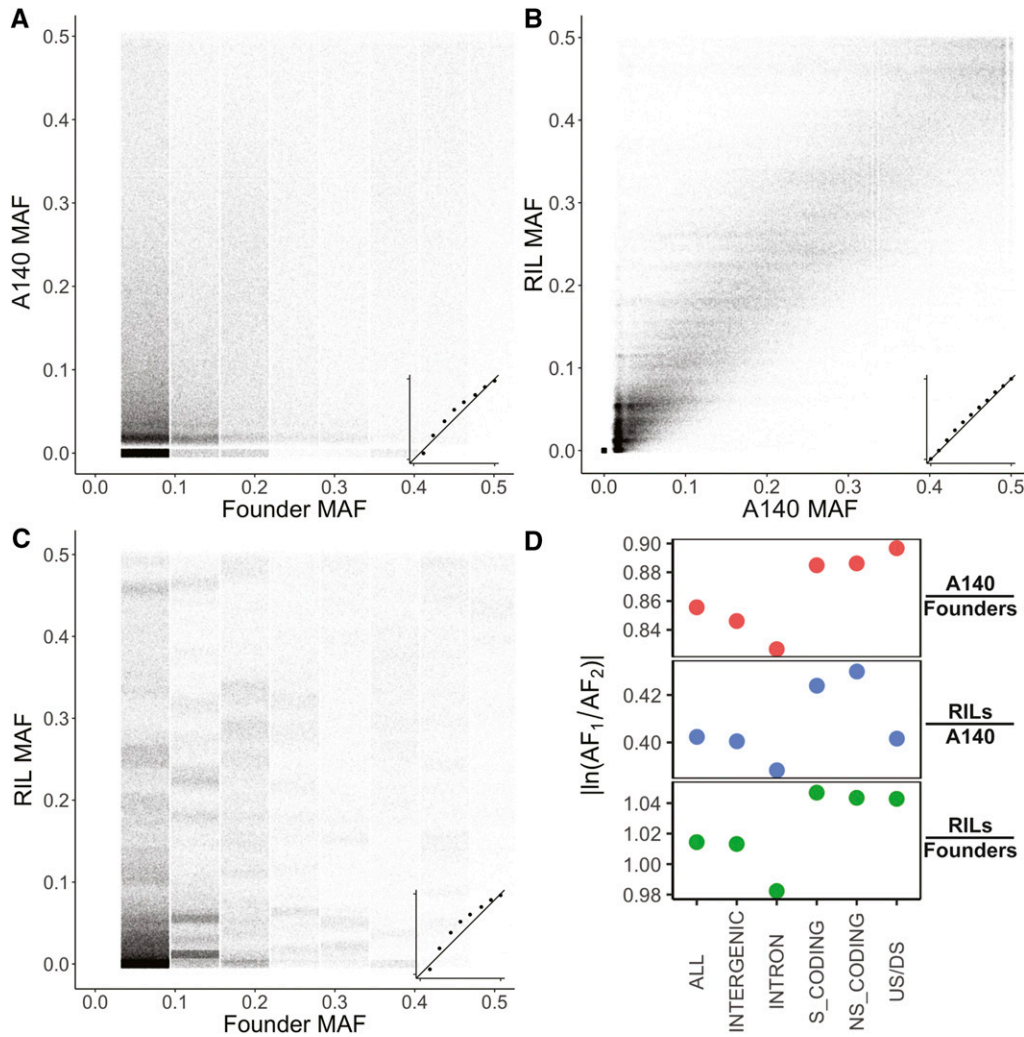
Figure 2 MAFs of founders and the outbred A140 population (A), A140 and RILs [inbreeding only for the A140 RILs, further adaptation then inbreeding for G50 RILs; (B)], and founders against all RILs (C). Insets show frequency quantiles. Changes in major/minor class across contrasts are ignored (among these cases, unfolded frequencies for just 3699 sites differ by > 50% for founders *vs.* RILs). (D) Change in allele frequency (absolute log ratios) for the same contrasts by functional class: intronic, synonymous and nonsynonymous, putative regulatory variation (US/DS; ≤ 200 bp from an annotated transcript or N2 pseudogene), or intergenic (none of the above). Points are mean values (diameter exceeds the SE). DS, downstream; MAF, minor allele frequency; RILs, recombinant inbred lines; US, upstream.

bottlenecks during serial intercrossing and population expansion, gives ample opportunity for loss of diversity through drift and selection. Fixation of N2 alleles at one X chromosome locus, spanning the known major effect behavioral locus *npr-1* (de Bono and Bargmann 1998; Gloria-Soria and Azevedo 2008; McGrath *et al.* 2009; Reddy *et al.* 2009; Bendesky *et al.* 2011; Andersen *et al.* 2014), during establishment of the A140 population has been documented with a coarse marker set (Teotónio *et al.* 2012). More broadly, the outbred A140 population showed nonnegligible departure from the founders, with 32,244 alleles lost (unseen in both the A140 and RILs, 26,593 of these being founder singletons; Figure 2). Subsequent change during the inbreeding (and further adaptation) stages to generate RILs was more restricted, with an additional 3171 alleles lost (2542 of these at < 10% frequency in both founders and the A140). However, importantly, the physical distribution of allelic loss is relatively restricted: at least one marker is segregating in the CeMEE RILs at > 5% MAF within 95.5% of 20 kb segments across the genome (97.2% of autosomal segments; for reference, protein coding genes are spaced just under 5 kb apart on average in the 100 Mb *C. elegans* N2 genome).

Analysis of differentiation across variant functional classes showed large departures in frequency for coding variation (synonymous and nonsynonymous) and the smallest for intronic variation (Figure 3D). Putative regulatory variation was highly variable across experimental phases, being the most dynamic class during the funnel intercross and initial adaptation (founders to A140) stage, but changing less than the mean value for all classes for generations between the A140 and RILs. This pattern was observed across all of the subpanels that make up the CeMEE (data not shown), notably the A140 RILs that differ from the outbred A140 by only inbreeding, suggesting differential dominance of coding and regulatory variation (Wray 2007; Gruber *et al.* 2012). Without sequence data for the outbred CA50, GA50, GM50, or GT50 populations, we cannot assess the impact of inbreeding on the fixation of alleles more generally. These effects are expected to depend on reproductive mode and selection (Charlesworth and Wright 2001; Morran *et al.* 2009; Chelo and Teotónio 2013; Chelo *et al.* 2014; Kamran-Disfani and Agrawal 2014).

Local LD, while nonuniform among chromosomes, tends to decay relatively rapidly, approaching background levels
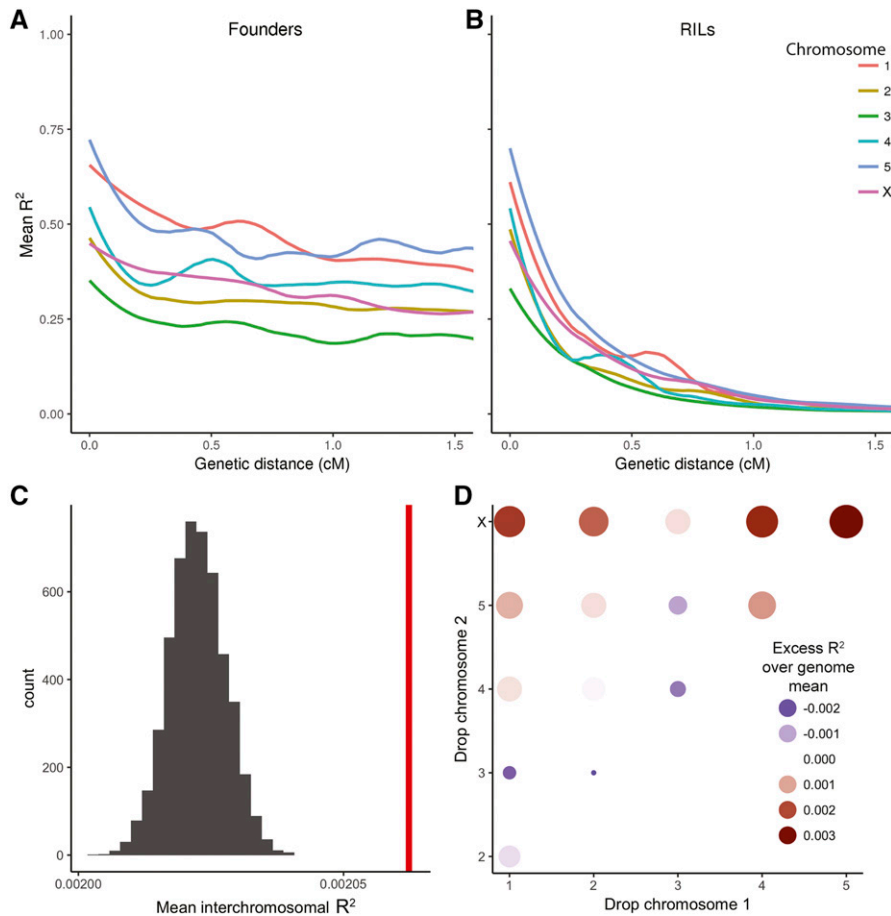
Figure 3 Linkage disequilibrium in founders (A) and all CeMEE RILs [(B); F_2 genetic map distance, LOESS fit to mean r^2]. (C) Interchromosomal structure is weak but significant. Observed mean r^2 between all chromosomes (red vertical bar) plotted against the null distribution from permutations randomizing lines across chromosomes (within subpanels to exclude effects of population structure). (D) Permutations dropping pairs of chromosomes implicate X–autosome interactions. Color and size are (redundantly) scaled by enrichment over the null distribution (95% percentile), relative to the genome-wide mean value. CeMEE, *C. elegans* multiparental experimental evolution panel; LOESS, LOcal regrESSion; RILs, recombinant inbred lines.

by 0.5 cM (F_2 map scale) on average (Figure 3 and Figure S3 in File S22). Disequilibrium between pairs of loci on different chromosomes is, as expected, very weak (r^2 0.99, 0.95 quantiles = 0.538, 0.051 within chromosomes *vs.* 0.037, and 0.022 across chromosomes), with one prominent exception. At a single pair of loci on chromosomes II and III, we observe $r^2 > 0.5$ [II: 2,284,322, which tags an intact MARINER5 transposon (WBTransposon00000128) that harbors an expressed miRNA in the N2 reference, and III: 1,354,894–1,425,217, a broad region of mostly unannotated genes]. The maximum interchromosomal r^2 for all other pairs is ≤ 0.27. Genotypes in repulsion phase are rare across these regions in the RILs ($P < 10^{-70}$, Fisher's Exact Test), absent in the founders, and present in only 1 of 124 wild isolates surveyed with unambiguous variant calls in these regions (CeNDR). This suggests the presence of at least one two-locus incompatibility exposed by inbreeding or, perhaps more likely given the uncertainties of reference-based genotyping, a transposon-mediated II–III transposition polymorphism among founders. Three founders contribute the chromosome II nonreference haplotype, but extremely poor read mapping in this region for these and other isolates, consistent with high local divergence as well as potential structural variation, means our short-read data are not informative in resolving these alternatives.

To better quantify the extent of subtle interchromosomal structure in the CeMEE, we compared the observed correlations among chromosomes to values from permutations, shuffling lines within subpanels, among chromosomes (Figure 3). The observed mean value for the genome, while extremely low, is higher than the maximum value of the null distribution ($P < 2 \times 10^{-4}$ from 5000 permutations), indicating the presence of extensive weak interactions. Further permutations dropping single or pairs of chromosomes showed that interactions between autosomes and the X chromosome contribute disproportionately (although all pairwise combinations exceeded the respective null maxima).

Founder haplotype blocks and genetic map expansion

The CeMEE panel is highly recombined and any simple, large-effect incompatibilities between founders are likely to have been purged early on in experimental evolution. For example, a haplotype containing *peel-1* and *zeel-1*, a known incompatibility locus that segregates among the founders on the left arm of chromosome I (Seidel *et al.* 2008, 2011), is fixed in the RILs (Figure 4A). Cases such as this are best appreciated when the mosaic of founder haplotypes that comprise each RIL genome is inferred.

Founder haplotypes in the RILs were reconstructed with the multiparent HMM framework RABBIT (Zheng *et al.* 2015), assigning 96.9% of markers to a single founder haplotype at posterior probability > 0.2 (84.2% > 0.5; median value across lines; haplotype sharing in the 16 founders
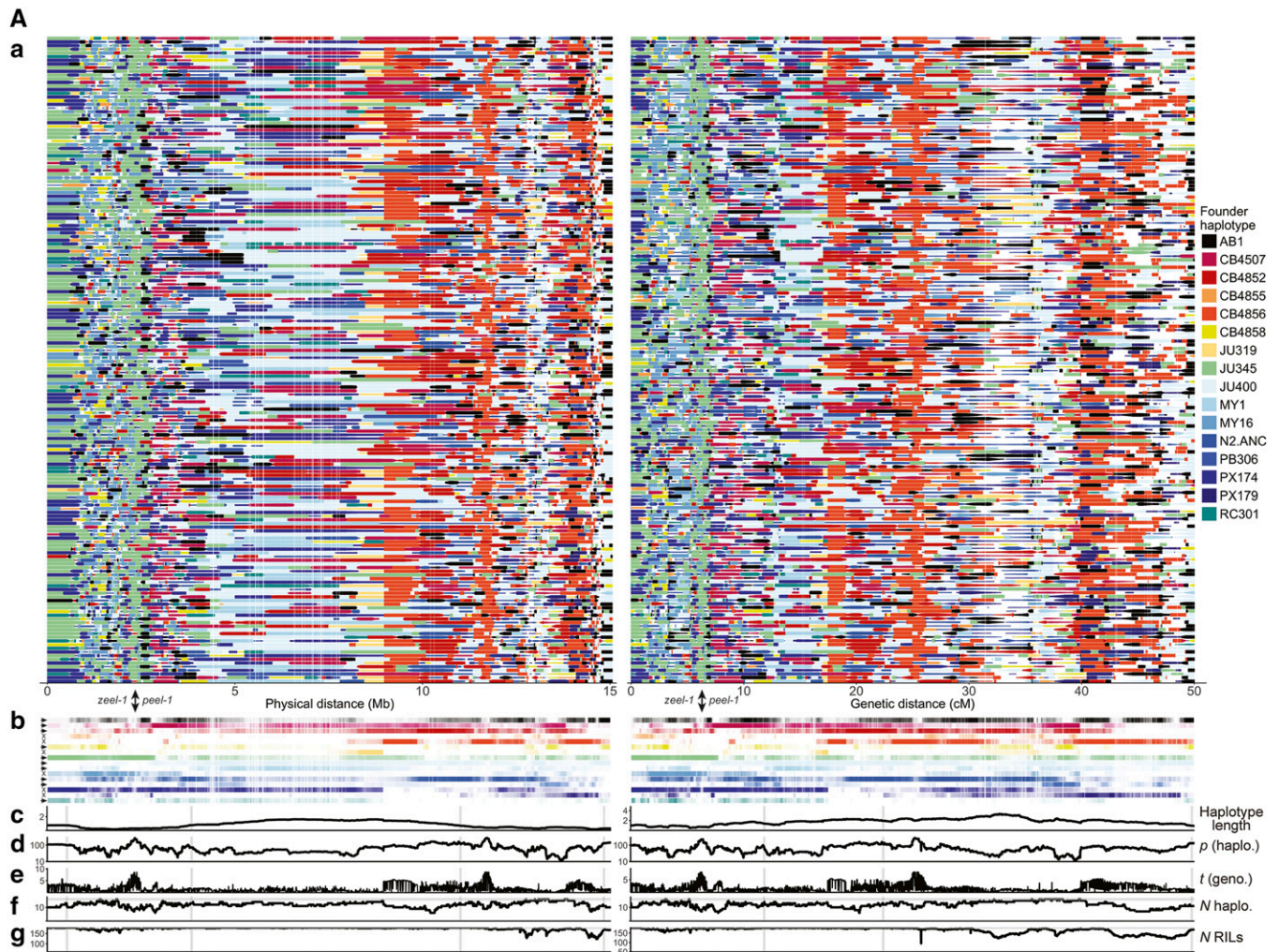
Figure 4 A140 RIL founder haplotype reconstruction and structure for chromosomes I (A), IV (B), and *X* (C). Founder haplotypes in physical and genetic distances (in subpanels labeled "a"). Each plotted point is a marker, with its size scaled by posterior probability (minimum 0.2; regions of low marker density are visible as vertical white swathes through RIL haplotypes). Founder contributions are summarized below (in subpanels labeled "b"). Loci discussed in the text are indicated: the *zeel-1/peel-1* incompatibility on the left arm of chromosome I (haplotype compatibility group), either experimentally tested in Seidel *et al.* (2008) or determined here from genotype data, is indicated below as an arrowhead for Bristol (N2) or an "x" for Hawaii (CB4856); extreme haplotype differentiation within a piRNA cluster on the right arm and tip of chromosome IV; and the fixation of N2/CB4507 haplotypes over a large region of the *X* chromosome left arm spanning *npr-1*, alleles of which have pleiotropic effects on behavior and laboratory adaptation (de Bono and Bargmann 1998; Gloria-Soria and Azevedo 2008; McGrath *et al.* 2009; Andersen *et al.* 2014). Subpanels "c–g" show summary statistics evaluated at 5 kb or 0.01 cM resolution, with vertical scales for each metric fixed across chromosomes, and the positions of recombination rate boundaries inferred for the N2 $\times$ CB4856 RIAILs (Rockman and Kruglyak 2009) indicated with shaded bars. Haplotype length: mean length extending from the focal position (in subpanels labeled "c"). *P* (haplo.): test of reconstructed founder haplotype proportions, relative to expectation based on reconstruction frequency from G_{150} simulations ($-log_{10}(P)$ from a χ^2 goodness-of-fit test) (in subpanels labeled "d"). *t* (geno.): change in allele frequency from the founders (absolute value of Welch's *t* statistic for founder *vs.* RIL genotype counts) (in subpanels labeled "e"). *N* haplo.: the number of unique founder haplotypes detected at each position, with the maximum value of 16 indicated (in subpanels labeled "f"). *N* RILs: the number of RIL haplotypes reconstructed at each interval (>0.2 posterior probability), with the maximum value of 178 indicated (in subpanels labeled "g"). RIAIL, recombinant inbred advanced intercross line; RIL, recombinant inbred line.

means that unambiguous assignment to a single founder is not always possible). For illustration purposes, a summary of reconstructed haplotypes for the A140 RILs on chromosomes I, IV, and *X* are shown in Figure 4, at both physical and genetic scales, to make the differences between these units plain. The observed recombination landscapes generally recapitulate those inferred from classical linkage mapping studies and from the N2/CB4856 RIAIL cross (Rockman and Kruglyak 2009), with the recombination rate high in chromosome arms and low in centers. With the additional map expansion gained

here (see below), we note that suppression of recombination is clearly strong, but not complete, within subtelomeric regions (see, for example, the exceptionally large right tip of chromosome *X*, spanning almost 2 Mb, in Figure 4C).

In general, founder haplotype diversity among RILs remains high: across reconstructed intervals, the median number of haplotypes observed in >1 RIL is 12 (posterior probability >0.5). Contributions clearly vary from equality, with lines most divergent from the N2 reference (CB4856 and MY16) overrepresented and lines more similar to the reference underrepresented
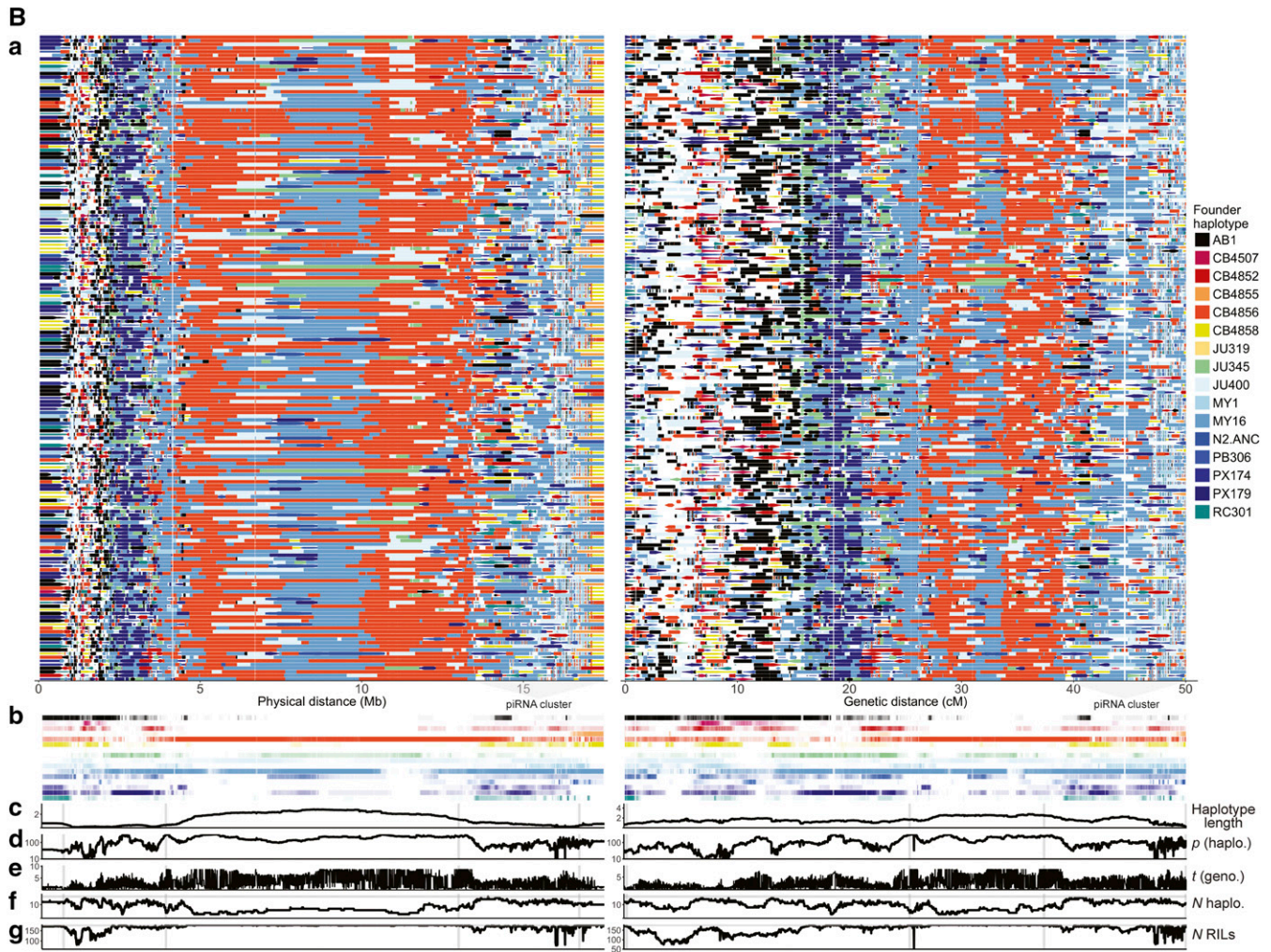
Figure 4 Continued.

(with the exception of the large region of chromosome *X*, spanning *npr-1*, which is largely fixed for N2/CB4507 alleles (Figure 4C). To examine whether these biases were merely technical, and establish expectations for reconstruction completeness and resolution in the presence of haplotype redundancy, we simulated genomes of varying pedigree length (up to 300 generations). As expected, reconstruction was hampered by increasing recombination, and by ambiguity between similar founders (Figure S4 in File S22). However, bias toward divergent haplotypes was not observed in the reconstruction simulations, suggesting that the overrepresentation of CB4856 and MY16 may be due to selection, notably for long haplotypes across the central domain of chromosome IV (Figure 5). Reconstruction completeness for the A140 RILs is generally in line with expectations for a pedigree of 150 generations. Clear exceptions are chromosome IV, where we recover more than expected under random sampling, and chromosome V, where we recover less. Haplotype lengths from simulated reconstructions showed that we progressively underestimate recombination breakpoints due to imperfect resolution of small haplotypes (Figure S4C in File S22).

The relationship between known generation and estimated realized map expansion from reconstruction simulations allows prediction of the number of effective generations of outcrossing. Across the five CeMEE subpanels, mean autosomal generation ranges from 227 (GM monoecious RILs) to 284 [control adapted (CA) androdioecious lines], with a weighted average of 260 for the panel as a whole (Figure S5 in File S22). Estimated genetic map expansion is highly variable across chromosomes: IV appears to be exceptionally recombinant in all subpanels, with expansion more than twice that of chromosomes I–III, largely due to a high-frequency, highly structured haplotype on the far-right arm and tip (Figure 4B (panel a)). This region spans one of the two large *C. elegans* piRNA clusters (Ruby *et al.* 2006), which encodes > 15,000 piRNA transcripts, interspersed with active transposons and protein-coding genes. Several trivial explanations for the unusual apparent expansion, such as elevated genotyping error rate or founder haplotype ambiguity, or distortions in the N2/CB4856 genetic map use to condition reconstruction probabilities, are not supported (data not shown), although the extent of large-scale structural variation among founders in this region [with the exception of CB4856, which does not show unusual levels of SNP or
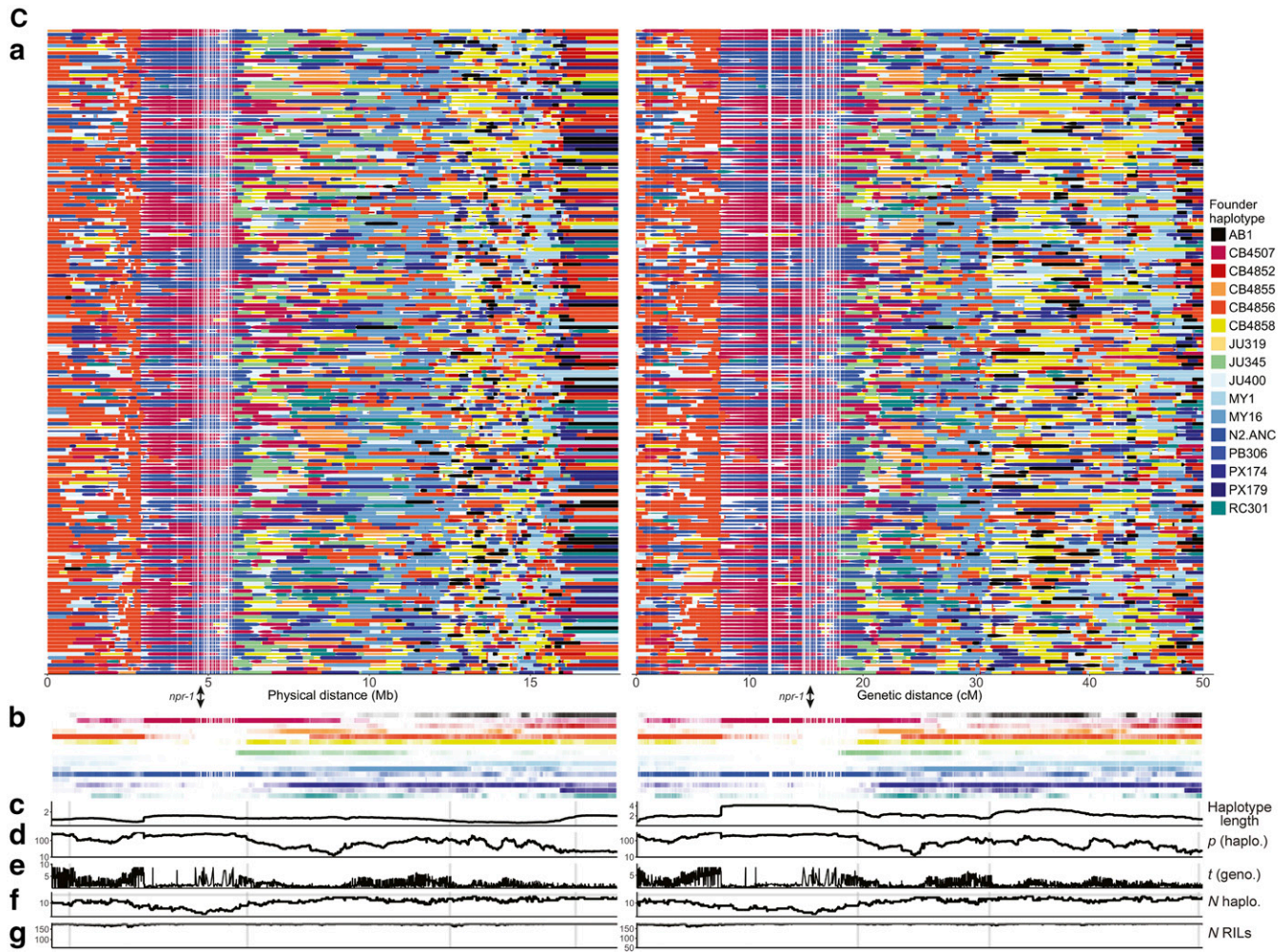
Figure 4 Continued.

copy number variation in the assembly of Thompson *et al.* (2015)] is unknown. Potential technical artifacts aside, the locus may represent a hitherto undetected recombination hotspot (whether through attraction or suppression of observed recombination elsewhere on the chromosome), a site of rampant gene conversion, or the focus of early and sustained selection during the initial intercross phase [potentially epistatic in nature, see Neher and Shraiman (2009)]. We previously proposed that evolution of this region may have involved a recombination rate modifier (through gene conversion) during the first 140 generations of experimental evolution to explain the observed excess haplotype diversity [see *Results and Discussion* and Figure S4 and Figure S5 of Chelo and Teotónio (2013)]. In contrast, chromosome V, which has been the focus of a recent large-scale selective sweep (Andersen *et al.* 2012), shows more variable expansion across subpanels suggestive of ongoing evolution (Figure S5 in File S22).

Population stratification

We examined additional genetic structure in the CeMEE RIL panel stemming from the presence of distinct subpanels of RILs that vary in experimental evolution histories. In the context of QTL mapping, this structure can represent nuisance variation that can bias estimates of heritability if unknown factors covary with the trait of interest, structure that is causally associated with a trait, or noncausal structure due solely to population stratification (or any combination of these factors).

To gauge the extent of population stratification, we compared the results of supervised and unsupervised DAPC (Jombart 2008), which partitions within and between group variation, using either known or inferred populations, based on linear combinations of PCs. By selection of discriminant functions that best predict known CeMEE subpanel membership, it is clear that the varied evolutionary history has, unsurprisingly, generated significant genetic structure. The number of PCs selected by cross-validation that best predicts population membership is 40, which together explain 25% of the variance (though only a fraction of these components are significantly associated, considered singly or in pairs, see *Materials and Methods*). Unsupervised DAPC, which infers groups based on variance minimization and model penalization criteria (k-means clustering, BIC), selected 5–8 clusters that best explain the data (Δ BIC < 1 over this range), although the rate of successful assignment was low (*e.g.*, 36% at the true value of $k = 5$). Together, these results
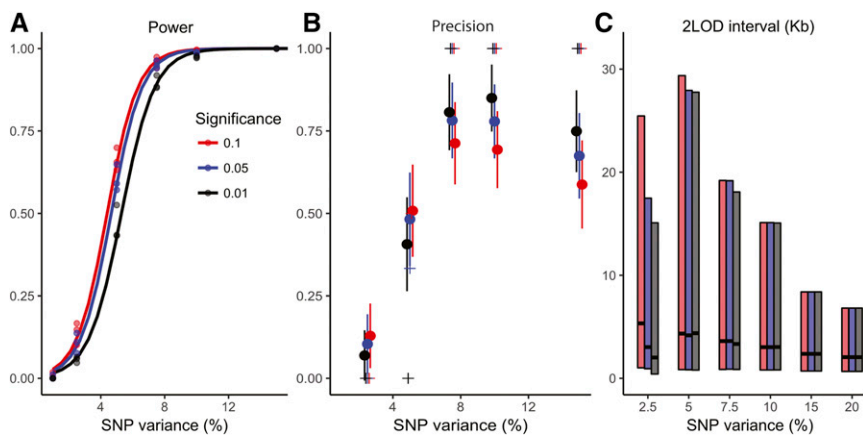
Figure 5 Additive QTL mapping simulations. Detection power (A), precision (B), and resolution (C) (2-LOD drop interval size for detected QTL) from single QTL simulations for the full mapping panel of 507 lines, as a function of detection threshold (significance at 0.01, 0.05, and 0.1) and phenotypic variance explained by the simulated QTL. Total heritability of simulated phenotypes is twice that of the focal QTL, with the polygenic contribution spread over 10, 100, or 1000 background markers [plotted in (A) and combined in (B and C)]. In (B) points are mean ± SE. Mean precision declines with SNP variance at high levels as chance associations reach significance, although the median value (+ symbols) is 1.0 at 5% significance for variants that explain ≥ 7.5% of trait variance. In (C) boxes span the interquartile range, with the median value indicated with a black bar.

suggest that genetic structure within, as well as between subpanels, is of comparable magnitude.

Heritability and predictability of fitness-proximal traits

We measured two traits that are important components of fitness as defined under our experimental evolution scheme, the fertility and size of young adult hermaphrodites, and thus represent challenging case studies for mapping of complex traits in the panel (Poullet *et al.* 2016). The traits are correlated (Figure S2 in File S22) and vary extensively in the CeMEE RILs: hermaphrodite fertility varies more than fivefold and size varies more than threefold (Figure 6).

Repeatability, genomic heritability, and prediction: RIL repeatability [an upper bound on broad sense heritability, H^2, under certain assumptions (Dohm 2002)] for both traits was relatively high −0.76 for fertility and 0.80 for size. However, additive genetic similarity based on the probability of allele sharing at all markers, equally weighted, gave genomic heritability estimates not significantly different to 0 (LRT; data not shown). This suggested that genome-wide genotypic similarity is poorly correlated with causal variation for these traits, potentially due to variable LD or epistatic cancellation (Lachowiec *et al.* 2015). Thus, we examined alternative measures of genetic similarity to address the apparent lack of additive genomic heritability, varying assumptions about the relationship of MAF, LD, and effect sizes, and the distribution of causal variation among chromosome and recombination rate domains. Model predictive power (r^2) was compared by leave-one-out cross-validation (see *Materials and Methods* and File S22).

Heritability estimates and prediction accuracy are summarized in Table 1, comparing the simplest models—additive (A) only, or additive + additive-by-additive (A^2) genetic covariance at the genome level—and the most predictive models for each trait. Given relatively high variance in relatedness, we are powered to detect large differences in additive heritabilities despite modest sample sizes for analysis of this kind, although the differences between individual models are generally minor. For fertility, with just 227 lines,

we have 50% power at a significance level of 0.05 to reject $h^2 = 0$ if $h^2 = 0.38$, and > 95% power at our estimate of H^2 (and marginally better power for size), based on the best-performing measure of additive similarity for each trait (Visscher *et al.* 2014). Low power to detect statistical epistasis is unavoidable given the multiplicative scaling of variance in genetic similarity.

While phenotype prediction accuracy is generally poor, some broad trends are apparent in the ranking. Additive heritability based on LD-weighted markers was relatively high for size (0.58), though less so for fertility (0.24). However, in neither case was additive similarity alone the best predictor of phenotype. Nine of the top 10 models for fertility incorporated epistasis in some form, with the best of these giving 57% improvement over the best additive model. For size, the advantage was less clear: three of the top four models included epistasis, although the performance differential between the best epistatic and additive models was only 3%.

Notably, partitioning of the genome based on recombination rate domains performed well for both traits, and was the preferred model for fertility. In general, scaling the expected relationship between allele frequency and effect size (α; see *Materials and Methods*) had weak effects on prediction. However, within models (varying only α), negative values were generally preferred for size (rarer alleles having larger effects) and positive for fertility, suggesting that the frequency spectrum of causal variants for the two traits varies in the RILs. Nevertheless, the additive genetic correlation between traits was generally high (*e.g.*, $r_G = 0.54$ for genetic similarity unweighted by allele frequency).

Effects of population stratification on heritability estimation

Given the stratified nature of the CeMEE panel, we tested for effects on heritability estimation in three ways. First, we estimated heritability for individual subpanels (best additive models only). Although highly uncertain given the very small sample sizes, estimates were positive for two of the three subpanels for adult body size and for both of two subpanels
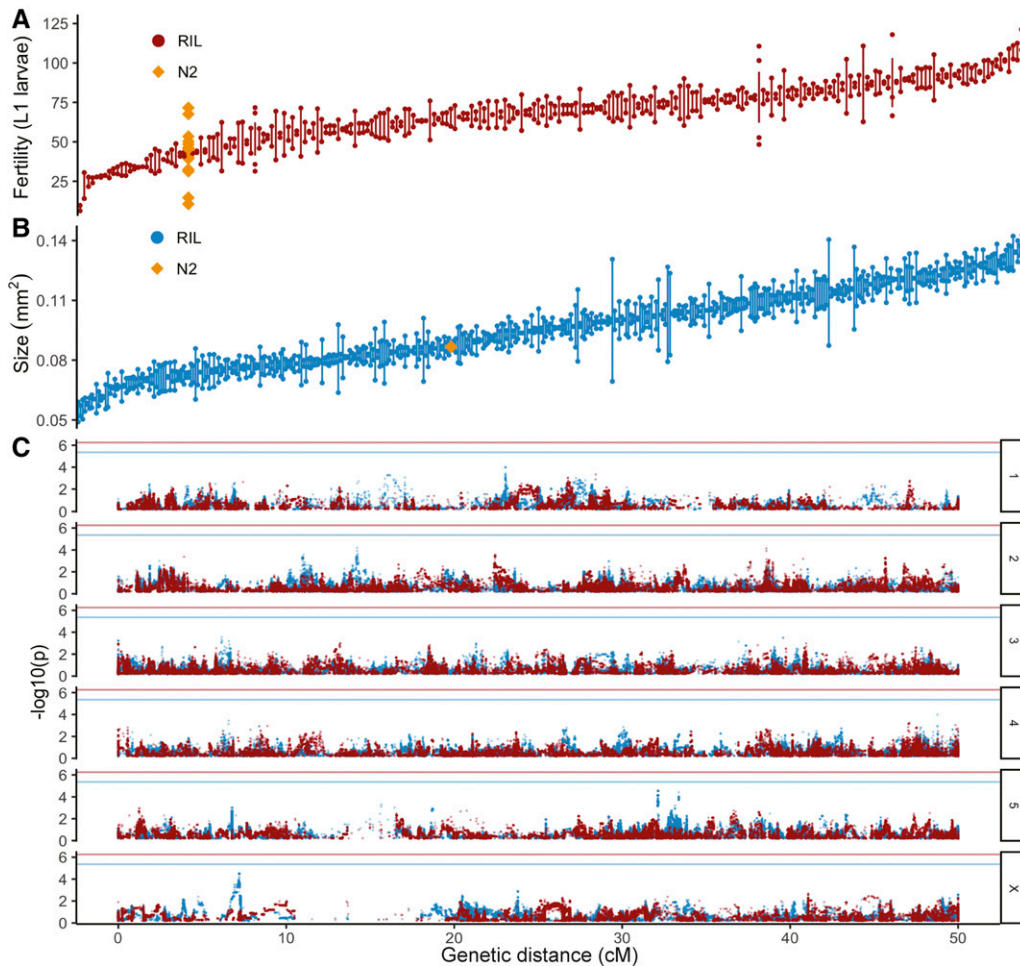
Figure 6 One-dimensional GWAS. (A and B) Trait value distributions across RILs (replicate means; bars show data range or the SE for samples with > 2 replicates). Values for the reference N2 strain are shown for comparison. Note that values are raw replicate means on the original scale, and so include all sources of technical variation (unlike model coefficients used for mapping). (C) Single-marker association results for fertility and adult body size (colors as above). $\alpha = 0.1$ thresholds = 4.38×10^{-6} and 5.57×10^{-7} for size and fertility, with minimum observed P-values of 2.8×10^{-5} and 7.23×10^{-5}, respectively. GWAS, genome-wide association study; RIL, recombinant inbred line.

tested ($n > 50$) for fertility, spanning the reported values for all lines.

Second, we estimated within-subpanel heritability by fitting population means as covariates (best $\mathbf{A}$ and $\mathbf{A}+\mathbf{A}^2$ models). For adult body size, where GA50 RILs are significantly larger than other panels, this reduced estimated heritability to 0.15 (**A**, from 0.58) and 0.38 ($\mathbf{A}+\mathbf{A}^2$, with $\mathbf{A}^2$=0.30). Fertility, for which trait values vary only weakly with subpanel, was largely unchanged at 0.45 (**A**) and the (unreasonably high and uncertain) estimate of 1.44 (SD 0.75) for $\mathbf{A}+\mathbf{A}^2$, with a dominant contribution from epistasis.

Third, we applied the method of Yang *et al.* (2011), developed for unrelated human populations, which compares the sum of heritabilities estimated for single chromosomes to that of a model fitting all chromosomes jointly. In the former case, genetic correlations across chromosomes due to population structure will result in $\sum h^2_{C(single)} > h^2_C$, since the genotype of one chromosome will be predictive of that of others, while fitting all chromosomes jointly gives independent conditional estimates. The reasonable underlying assumptions are that structure is more significant between than within populations, and is not causally associated with phenotypic variance, although the latter might not hold for fitness-proximal traits. Comparing the sum of heritability estimates from samples of half the chromosomes ($\sum h^2/2$) to that

from all chromosomes (additive similarity only), results suggested that stratification may contribute significantly to our estimates for size, with mean $\sum h^2/2 = 0.72$ (contributing 20% of the total given $h^2 = 0.60$ for a joint chromosome model), but not for fertility (mean $\sum h^2/2 < h^2$). Fitting ≤ 80 PCs as covariates for size failed to bring this ratio to equality, but progressively eroded the heritability estimate (minimum 10% "inflation" for 80 PCs, $h^2 = 0.30$), while fitting three DAPCs (based on the top 40 PCs) fully accounted for the difference (mean $\sum h^2/2 = h^2 = 0.39$). Notably, however, performing the same analysis within subpanels gave a similar level of inflation for size within the largest group of RILs (28%), suggesting that structure not associated with subpanel is also influential.

The above analyses lead us to conclude that the results presented in Table 1 for fertility are robust, while those for adult size are somewhat less so. However, the extent of inflation is unlikely to be as severe as indicated by disjoint genome partitioning, and no covariates were fit for subsequent analyses.

GWAS

QTL mapping power and precision: We first explored general characteristics of the CeMEE panel relevant to mapping quantitative traits. Association tests were carried out by LMM on simulated phenotypes, varying the effect size of causal

Table 1 Genomic heritability estimates

Trait	GSM	α	r^2	h^2 (SD)	LR
Size	A	-0.5	0.073	A 0.58 (0.14)	11
	A + A^2	**-0.5**	**0.093**	**A 0.57 (0.15)**	**11**
				A^2 0.21 (0.51)	
Fertility	A	1	0.012	A 0.24 (0.24)	0
	A + A^2	1	0.029	A 0.36 (0.21)	3
				A^2 1.24 (0.87)	
	(A + A^2)$_{rec}$	**1**	**0.064**	**A$_{arm}$ 0.44 (0.18)**	**7**
				A$_{cen}$ 0.02 (0.07)	

Results are shown for additive (A) and additive-by-additive (A^2) GSM, and for the most predictive model tested (if neither of the above), shown in bold. α is the scaling parameter from Speed *et al.* (2012), which determines the effect size expectation for markers as a function of allele frequency, where 0 is unweighted and smaller values assign greater weight to rare alleles (see Equations 1 and 2). Unconstrained REML estimates and SD are shown for components > 0 at convergence. *LR* is improvement over the null model (likelihood ratio). A + A$^2_{rec}$ is additive and additive-by-additive similarity defined at the level of recombination rate domains (tips, arms and centers, merged across chromosomes). See Equation S4 and associated discussion in File S22. GSM, genetic similarity matrices; LR, log likelihood ratios; REML, restricted/residual maximum likelihood.

variation and the degree of polygenicity (see *Materials and Methods*). For an allele explaining > 4.7% of the phenotypic variance, the power to detect the association is > 50% (permutation 5% significance threshold of $P < 1.62 \times 10^{-6}$), with precision (% true positives of all positives) > 50%, (Figure 5). When detected, the median QTL support interval (a drop in LOD score of 2) spans < 10 kb for variants explaining > 2.5% of trait variance. Given an average gene size of ~5 kb in *C. elegans* N2, including intergenic sequence, the CeMEE reaches subgenic resolution for alleles of large effect (> 10%), yielding high mapping precision (Figure 5). We note that our simulations are unbiased with respect to chromosomal location. If causal variation is enriched on the highly recombinant arms, these estimates are likely to be conservative.

1D mapping of fertility and size: We tested for single-marker effects on size and fertility by LMM, controlling for genome-wide relatedness using the most predictive LD-weighted additive GSM for each trait (see above). Based on permutation thresholds, no single markers were significant (Figure 6). For size, *P*-values were moderately inflated at the high end, but they were strongly deflated for fertility, consistent with model misspecification. Results were largely independent of the method used to define similarity or, for fertility, whether correction for relatedness was applied at all (Figure S6 in File S22). LD score regression, a related approach that explicitly assumes an infinitesimal architecture (Bulik-Sullivan *et al.* 2015), gave further support for extensive polygenicity with effects distributed across the genome (again, mostly clearly for fertility; Figure S7 in File S22). Given significant heritabilities for both traits and the results of GWAS simulations, the absence of individually significant associations suggests architectures comprising many variants with additive effects explaining < 5% of the phenotypic variance.

2D mapping of additive-by-additive interactions: Given suggestive evidence for epistasis from variance decomposition and a lack of individually significant additive effects by 1D

mapping, we sought to identify interactions among markers underlying variance in size and fertility. At a significance level of $\alpha = 0.1$, we detected eight interactions (between 13 loci) for fertility and six (12 loci) for size, with modest marginal additive effects (Figure 7; median value of the minimum single-locus statistics per pair $P = 1.7 \times 10^{-4}$ for fertility and $P = 2.6 \times 10^{-5}$ for size). The variance explained by each pair, considered individually, is high: 12–15% for fertility and ~8% for size. The joint additive effects of these markers explain 9.1 and 9.5% of the phenotypic variances, respectively, while a full interaction model explains 38 and 32%.

By summing LRs in 1D space to test for polygenic epistasis, we also detected 31 unique markers with excess (1:many) interchromosomal interactions for fertility ($\alpha = 0.01$), and 77 for size Figure 7). A minority of these sites were also significant in single-pair tests (three for fertility and four for size). Notably, loci on chromosome IV in modest linkage (marker positions 1,888,439, 1,894,021, and 1,914,315 Mb) are involved in individually significant interactions of opposite effect with chromosomes II, III, and IV explaining 17% of the variance in fertility, and IV: 1,914,315 also shows a significant excess of interactions with chromosome II ($n = 132$). IV: 1,914,315 is found within an intron of *egl-18* (encoding a GATA transcription factor), while the two other markers fall within the exceptionally large serial intergenic region between *egl-18* and *egl-4* (encoding a cyclic-GMP-dependent protein kinase thought to act in the TGF-β pathway). Both genes vary in coding and UTR sequence among the founders, and have numerous known phenotypes from classical induced mutations and RNA interference spanning the gamut of behavior, development, and reproduction. Their eponymous phenotype, egg-laying abnormal (Egl), is retention of oocytes and embryos, a phenotype selected during experimental evolution in which embryos were extracted each generation by bleaching (Poullet *et al.* 2016).

Conclusions

We have described the generation, characterization, and application of the first multiparental mapping panel for the model organism *C. elegans*. Drawing on effectively 260 generations of outcrossing at moderate population size during laboratory culture, full reference-based genome sequencing of the 16 inbred wild founders, and dense genotyping of the RILs, the CeMEE panel yields gene-level mapping resolution for additive alleles explaining $\geq$ 5% of trait variance. For traits such as gene expression, for which local variation typically accounts for upward of 20% (*e.g.*, Brem and Kruglyak 2005; Rockman *et al.* 2010; King *et al.* 2014), the majority of (detected) QTL intervals will dissect single genes.

Multiparental mapping resources combine the power of controlled crosses with levels of diversity potentially approaching those of wild populations. Of existing resources, the CeMEE is most similar in spirit to the mouse Collaborative Cross (Churchill *et al.* 2004, 2012; Philip *et al.* 2011; Threadgill and Churchill 2012). Microscopic nematodes and macroscopic mammals vary in many ways, and major conveniences of the *C. elegans* system
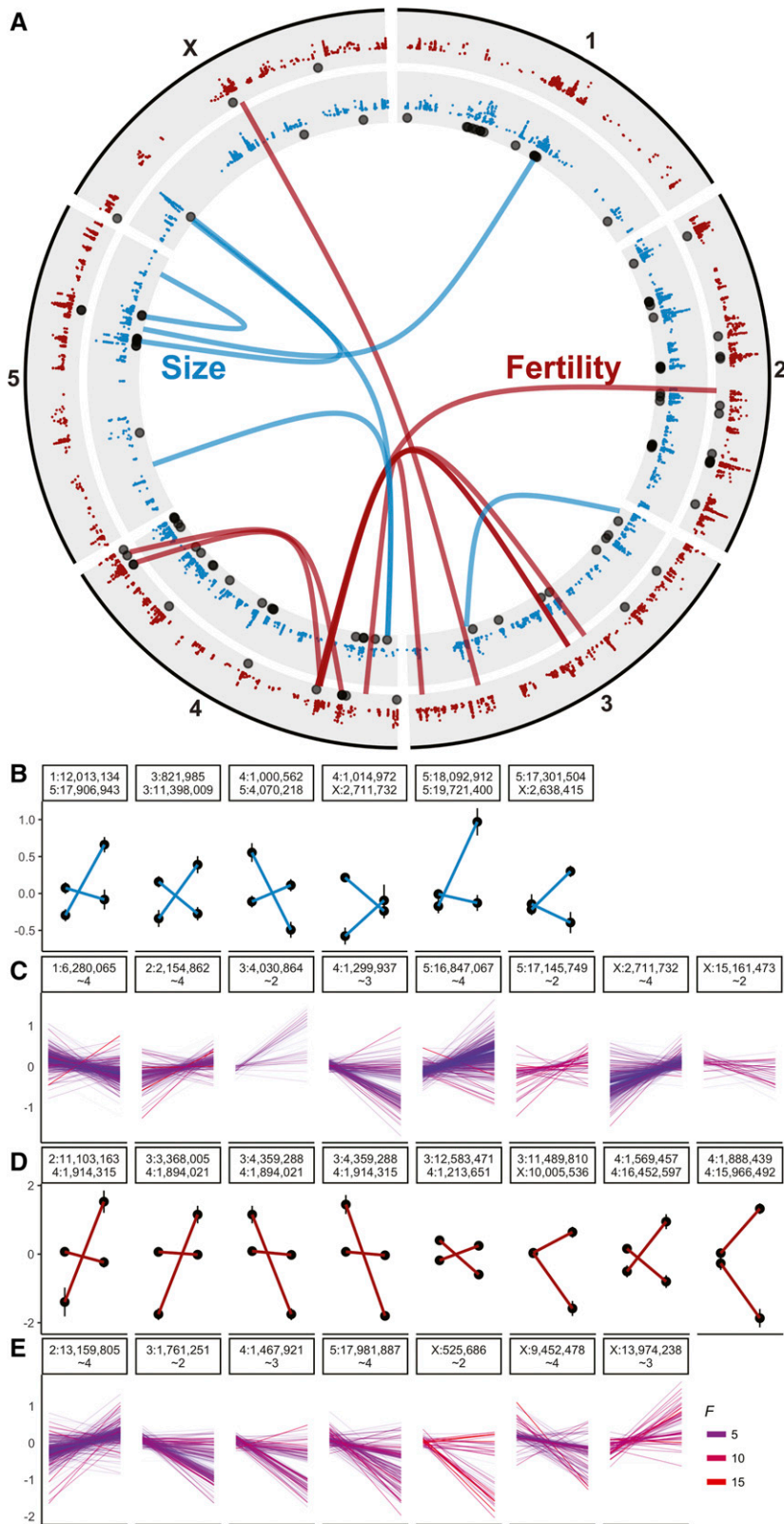
Figure 7 Strong-sign epistasis and highly polygenic interactions contribute to trait variance. (A) The distribution of significant interactions for fertility and size plotted by genetic distance. Pairwise interactions are linked over one-dimensional genome-wide association study test statistics ($-log10(P) > 1$) for each trait (size in blue and fertility in red). Markers with a significant excess of polygenic interactions are indicated with black points. These two-dimensional (2D) sum tests are one-to-many interactions between a single focal marker and all other markers on one other chromosome, with the sum of likelihood ratios significant under a null permutation model. (B and C) and (D and E) show genotype class means for significant size (and fertility) interactions: pairwise tests are in B and D (mean $\pm$ SE), and the individual pairwise tests that contribute to 2D sum tests are in C and E (mean values only). The marker with the highest summed likelihood ratio for each significant chromosome combination is shown. In C and D, line color and intensity is scaled by the interaction F statistic for each interaction according to the scale shown in (E).

such as parental zygosity, inbreeding by selfing without a high burden of lethality, advanced pedigree, cryopreservation, and small genome size, have facilitated the generation and characterization of a powerful and stable resource that is relatively representative of its founders. The comparison with Drosophila is perhaps more fair, where the Drosophila Genetic Reference Panel (DGRP) and Drosophila Synthetic Population Resource (DSPR) have been widely used to study the

genetic basis of complex organismal and molecular traits (King *et al.* 2012b, 2014; Swarup *et al.* 2013; Marriage *et al.* 2014; Dembeck *et al.* 2015; Huang *et al.* 2015; Najarro *et al.* 2015; Rohde *et al.* 2017). The DGRP GWAS panel preserves a sample of genetic variation from an ancestral population of large effective size in a relatively small number of RILs (185 with coefficient of coancestry < 0.25), and therefore is highly diverse, with an allele frequency spectrum skewed toward rare alleles. Although LD is practically absent, on average, it is highly variable due to sampling effects, exacerbated by large segregating inversions, leading to rarity disequilibria that can inflate false positive QTL (Houle and Márquez 2015). The DSPR, comprising two eight-parent, 50-generation RIL panels founded from 15 inbred isolates (> 1700 lines in total), achieves good power and decent resolution (84% power to detect QTL explaining 5% of trait variance in the pA panel by simulation, with a mean QTL interval of 1.5 cM, or around 630 kb; King *et al.* 2012a), with an allele frequency spectrum more favorable for QTL detection. The CeMEE panel, at its current size, performs similarly to the DSPR for mapping additive QTL (slightly lower power but considerably better resolution). Important considerations for mapping individual interactions and epistatic variance components are, respectively, the joint allele frequency spectrum and the degree of relatedness among lines (Young and Durbin 2014). In both these respects, the CeMEE is well placed relative to other model system panels.

The native androdioecious mating system of *C. elegans* and the ability to archive strains indefinitely confer almost microbial powers on a metazoan model. For one, the preservation of intermediate outbred populations means that the CeMEE is readily extensible, limited only by effective population sizes, and will benefit from a doubling of the number of RILs underway. However, RIL panels have several potential shortcomings. First, despite inbreeding during RIL construction, a nagging concern is residual heterozygosity (Barrière *et al.* 2009; Chelo *et al.* 2014) and the possibility of further evolution subsequent to initial characterization. While heterozygosity appears to be at a low level in the CeMEE RILs, on average, it is not absent (see *Materials and Methods*). However, importantly, given that lines are in stasis, the opportunity for segregation during further use is both limited and known. A second concern is the possibility of inbreeding depression, particularly for fitness-proximal traits. This applies predominantly to obligately outcrossing organisms (Barrière *et al.* 2009; Chelo *et al.* 2014), but it is also applicable to multiparental experimental evolution of *C. elegans*. As mentioned in the introduction, excess heterozygosity may have been maintained by epistatic overdominant selection during the initial stage of laboratory adaptation, and closely linked recessive deleterious alleles in repulsion could be maintained by balancing selection (Chelo *et al.* 2013, 2014). Assaying the F_1 progeny of nested crosses among RILs may be a useful approach to estimate (or avoid) the effects of inbreeding depression in the future (Long *et al.* 2014).

While reference-based genotyping will remain a practical necessity for some time yet, it leaves the contribution of certain classes of genetic variation uncertain, and can hamper variant calling due to mapping bias and erroneous alignments at copy number variants. The genome of only one wild isolate, the Hawaiian CB4856, has been assembled *de novo* to a high standard, revealing extensive divergence (Thompson *et al.* 2015). The ultimate goal of full genomes for all founders will yield both better accuracy in calculating genetic similarity and the ability to measure phenotypic effects of this recalcitrant variation. Similarly undetermined, given RIL genotyping by mostly low-coverage sequencing, is the extent of gene conversion and the fate of novel mutations during experimental evolution. With a mutation rate of around 1/genome/generation for SNPs, and more for multinucleotide mutations and copy number variation (Denver *et al.* 2004a,b, 2010; Seyfert *et al.* 2008; Phillips *et al.* 2009; Lipinski *et al.* 2011; Meier *et al.* 2014), the contribution of new mutations to trait variation in the RILs may well be nonnegligible. Theory suggests that fixation of adaptive mutations should not be significant during experimental evolution (Hill 1982; Caballero and Santiago 1995; Matuszewski *et al.* 2015), but empirical evidence is mixed (Estes 2004; Denver *et al.* 2010; Estes *et al.* 2011; Chelo *et al.* 2013). All of these factors would erode phenotype prediction accuracy, which, theoretically, should converge on H^2 given perfect genotyping of causal variation and appropriate description of genetic covariances (de los Campos *et al.* 2015).

Using subsets of the CeMEE panel, we outlined the genetics of two correlated traits associated with fitness. Variance in fitness-related traits, in particular, may be maintained despite consistent selection on additive variation through a number of processes, including stabilizing selection under a stable environment (Whitlock *et al.* 1995; Wolf *et al.* 2000; Barton and Keightley 2002; Phillips 2008; Hemani *et al.* 2013). Variance in size and fertility remains high, and we find evidence that this is due largely to abundant interactions. Notably for fertility, which is expected to be well aligned with fitness under the experimental evolution scheme, strong interactions among 13 alleles with weak to moderate marginal effects jointly explain over a third of the phenotypic variance. It is unclear why such large effect alleles are still segregating after extensive experimental evolution. Proximal explanations may lie in dominance, antagonistic pleiotropy, or in higher order properties of the genotype–phenotype map.

All pairwise interactions detected are instances of sign epistasis, where the directional effect of one allele is reversed in the presence of another. At least five of these represent the extreme form, reciprocal sign epistasis (the reversal is, to some extent at least, symmetric; Poelwijk *et al.* 2011). Sign epistasis, in particular, has important implications for a population's capacity to adapt, potentially creating rugged fitness landscapes and constraining exploration of them (Weinreich *et al.* 2005, 2013), and also for the repeatability of evolution, since the outcome of selection on the marginal additive effects of interacting alleles will be determined by their relative frequencies (Wright 1932; Whitlock *et al.* 1995; Phillips *et al.* 2000). Our tests for excess interactions among individually nonsignificant marker pairs additionally revealed many cases of highly

polygenic epistasis. While tests of this type have the unsatisfying property of leaving the identities of the interacting partners uncertain, they have the potential to combat the loss of power that comes with brute-force 2D testing (Crawford *et al.* 2016). The effects of these segregating interactions on the genotype–phenotype map can now be readily tested by replicated experimental evolution. In this context, it will be useful to determine the directional effects of epistasis during further evolution as a function of recombination and varied environment, a task for which the CeMEE is well suited.

Fertility and body size at reproduction show broad-sense heritabilities that are relatively high for fitness-proximal traits (Lynch and Walsh 1998), likely a consequence of novel genetic variation created in the multiparental cross and realignment of selection to novel laboratory environments. While all mapping panels are synthetic systems, the mixing of natural variation and experimental evolution represents a perturbation that may have some parallels, for example, with that of founder events or environmental change, which can reveal novel incompatibilities and promote further population differentiation (Cheverud and Routman 1996; Johnson 2000). As in other systems such as Arabidopsis, where similar resources exist (Weigel 2012) and epistasis for fitness-related traits has been found (*e.g.*, Malmberg *et al.* 2005; Simon *et al.* 2008), it will also be important to begin a comprehensive comparison of QTL for fitness traits in the CeMEE and natural populations— where linked selection coupled with predominant selfing and meta-population dynamics have generated limited structured genetic diversity (Rockman *et al.* 2010; Andersen *et al.* 2012; Cutter 2015)—and also with mutational variances obtained in mutation accumulation experiments (Baer *et al.* 2005; Baer 2008; Joyner-Matos *et al.* 2009). Such comparisons have the potential to provide insights into how QTL effects and frequencies are shaped in natural populations.

Acknowledgments

Author contributions: S.C., B.A., and H.T. conducted CeMEE panel derivation. A.P.-Q., D.D.R., I.C., P.A., L.M.N., and M.V.R. performed sequencing and genotyping. I.C., B.A., and A.C. conducted phenotyping; L.M.N., I.C., T.G., A.D., and B.S. analyzed data. L.M.N., M.V.R., and H.T. wrote the manuscript.

Literature Cited

Abney, M., 2015   Permutation testing in the presence of polygenic variation. Genet. Epidemiol. 39: 249–258.

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.

Andersen, E. C., J. S. Bloom, J. P. Gerke, and L. Kruglyak, 2014   A variant in the neuropeptide receptor *npr-1* is a major determinant of *Caenorhabditis elegans* growth and physiology. PLoS Genet. 10: e1004156.

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

Anderson, J. L., L. T. Morran, and P. C. Phillips, 2010   Outcrossing and the maintenance of males within *C. elegans* populations. J. Hered. 101(Suppl. 1): S62–S74.

Anderson, M., and C. Ter Braak, 2003   Permutation tests for multifactorial analysis of variance. J. Stat. Comput. Simul. 73: 85–113.

Baer, C. F., 2008   Quantifying the decanalizing effects of spontaneous mutations in rhabditid nematodes. Am. Nat. 172: 272–281.

Baer, C. F., F. Shaw, C. Steding, M. Baumgartner, A. Hawkins *et al.*, 2005   Comparative evolutionary genetics of spontaneous mutations affecting fitness in rhabditid nematodes. Proc. Natl. Acad. Sci. USA 102: 5785–5790.

Baldwin-Brown, J. G., A. D. Long, and K. R. Thornton, 2014   The power to detect quantitative trait loci using resequenced, experimentally evolved populations of diploid, sexual organisms. Mol. Biol. Evol. 31: 1040–1055.

Bandillo, N., C. Raghavan, P. A. Muyco, M. A. L. Sevilla, I. T. Lobina *et al.*, 2013   Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. Rice (N. Y.) 6: 11.

Barrière, A., and M.-A. Félix, 2005   Natural variation and population genetics of *Caenorhabditis elegans* (December 26, 2005), *WormBook*, ed. The *C. elegans* Research Community WormBook, doi/10.1895/wormbook.1.43.1, http://www.wormbook.org.

Barrière, A., and M.-A. Félix, 2007   Temporal dynamics and linkage disequilibrium in natural *Caenorhabditis elegans* populations. Genetics 176: 999–1011.

Barrière, A., S.-P. Yang, E. Pekarek, C. G. Thomas, E. S. Haag *et al.*, 2009   Detecting heterozygosity in shotgun genome assemblies: lessons from obligately outcrossing nematodes. Genome Res. 19: 470–480.

Barton, N. H., 2017   How does epistasis influence the response to selection? Heredity 118: 96–109.

Barton, N. H., and P. D. Keightley, 2002   Multifactorial genetics: understanding quantitative genetic variation. Nat. Rev. Genet. 3: 11–21.

Barton, N. H., and M. Turelli, 1991   Natural and sexual selection on many loci. Genetics 127: 229–255.

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.

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

Bonhoeffer, S., C. Chappey, N. T. Parkin, J. M. Whitcomb, and C. J. Petropoulos, 2004   Evidence for positive epistasis in HIV-1. Science 306: 1547–1550.

Brem, R. B., and L. Kruglyak, 2005   The landscape of genetic complexity across 5,700 gene expression traits in yeast. Proc. Natl. Acad. Sci. USA 102: 1572–1577.

Buckler, E. S., J. B. Holland, P. J. Bradbury, C. B. Acharya, P. J. Brown *et al.*, 2009   The genetic architecture of maize flowering time. Science 325: 714–718.

Bulik-Sullivan, B. K., P.-R. Loh, H. K. Finucane, S. Ripke, J. Yang et al., 2015 LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. 47: 291–295.

Bůžková, P., T. Lumley, and K. Rice, 2011 Permutation and parametric bootstrap tests for gene-gene and gene-environment interactions. Ann. Hum. Genet. 75: 36–45.

Caballero, A., and E. Santiago, 1995 Response to selection from new mutation and effective size of partially inbred populations. I. Theoretical results. Genet. Res. 66: 213–225.

Carlborg, Ö., L. Jacobsson, P. Ahgren, P. Siegel, and L. Andersson, 2006 Epistasis and the release of genetic variation during long-term selection. Nat. Genet. 38: 418–420.

Charlesworth, D., and S. I. Wright, 2001 Breeding systems and genome evolution. Curr. Opin. Genet. Dev. 11: 685–690.

Chelo, I. M., and H. Teotónio, 2013 The opportunity for balancing selection in experimental populations of Caenorhabditis elegans. Evolution 67: 142–156.

Chelo, I. M., J. Nédli, I. Gordo, and H. Teotónio, 2013 An experimental test on the probability of extinction of new genetic variants. Nat. Commun. 4: 2417.

Chelo, I. M., S. Carvalho, M. Roque, S. R. Proulx, and H. Teotónio, 2014 The genetic basis and experimental evolution of inbreeding depression in Caenorhabditis elegans. Heredity 112: 248–254.

Cheverud, J. M., and E. J. Routman, 1995 Epistasis and its contribution to genetic variance components. Genetics 139: 1455–1461.

Cheverud, J. M., and E. J. Routman, 1996 Epistasis as a source of increased additive genetic variance at population bottlenecks. Evolution 50: 1042.

Chirgwin, E., D. J. Marshall, C. M. Sgrò, and K. Monro, 2016 The other 96%: can neglected sources of fitness variation offer new insights into adaptation to global change? Evol. Appl. 10: 267–275.

Churchill, G. A., D. C. Airey, H. Allayee, J. M. Angel, A. D. Attie et al., 2004 The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat. Genet. 36: 1133–1137.

Churchill, G. A., D. M. Gatti, S. C. Munger, and K. L. Svenson, 2012 The diversity outbred mouse population. Mamm. Genome 23: 713–718.

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

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

Corbett-Detig, R. B., J. Zhou, A. G. Clark, D. L. Hartl, and J. F. Ayroles, 2013 Genetic incompatibilities are widespread within species. Nature 504: 135–137.

Corsi, A. K., B. Wightman, and M. Chalfie, 2015 A transparent window into biology: a primer on Caenorhabditis elegans. Genetics 200: 387–407.

Covarrubias-Pazaran, G., 2016 Genome-assisted prediction of quantitative traits using the R package sommer. PLoS One 11: e0156744.

Crawford, L., S. Mukherjee, and X. Zhou, 2016 Detecting epistasis in genome-wide association studies with the marginal EPIstasis test. bioRxiv DOI: 10.1101/066985 .

Cutter, A. D., 2004 Sperm-limited fecundity in nematodes: how many sperm are enough? Evolution 58: 651–655.

Cutter, A. D., 2006 Nucleotide polymorphism and linkage disequilibrium in wild populations of the partial selfer Caenorhabditis elegans. Genetics 172: 171–184.

Cutter, A. D., 2015 Caenorhabditis evolution in the wild. Bioessays 37: 983–995.

Cutter, A. D., A. Dey, and R. L. Murray, 2009 Evolution of the Caenorhabditis elegans genome. Mol. Biol. Evol. 26: 1199–1234.

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.

de los Campos, G., D. Sorensen, and D. Gianola, 2015 Genomic heritability: what is it? PLoS Genet. 11: e1005048.

Dembeck, L. M., K. Böröczky, W. Huang, C. Schal, R. R. H. Anholt et al., 2015 Genetic architecture of natural variation in cuticular hydrocarbon composition in Drosophila melanogaster. Elife 4: e09861.

Denver, D. R., K. Morris, A. Kewalramani, K. E. Harris, A. Chow et al., 2004a Abundance, distribution, and mutation rates of homopolymeric nucleotide runs in the genome of Caenorhabditis elegans. J. Mol. Evol. 58: 584–595.

Denver, D. R., K. Morris, M. Lynch, and W. K. Thomas, 2004b High mutation rate and predominance of insertions in the Caenorhabditis elegans nuclear genome. Nature 430: 679–682.

Denver, D. R., D. K. Howe, L. J. Wilhelm, C. A. Palmer, J. L. Anderson et al., 2010 Selective sweeps and parallel mutation in the adaptive recovery from deleterious mutation in Caenorhabditis elegans. Genome Res. 20: 1663–1671.

Diaz, S. A., and M. Viney, 2014 Genotypic-specific variance in Caenorhabditis elegans lifetime fecundity. Ecol. Evol. 4: 2058–2069.

Dohm, M. R., 2002 Repeatability estimates do not always set an upper limit to heritability. Funct. Ecol. 16: 273–280.

Dolgin, E. S., B. Charlesworth, S. E. Baird, and A. D. Cutter, 2007 Inbreeding and outbreeding depression in Caenorhabditis nematodes. Evolution 61: 1339–1352.

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

Duveau, F., and M.-A. Félix, 2012 Role of pleiotropy in the evolution of a cryptic developmental variation in Caenorhabditis elegans. PLoS Biol. 10: e1001230.

Estes, S., 2004 Mutation accumulation in populations of varying size: the distribution of mutational effects for fitness correlates in Caenorhabditis elegans. Genetics 166: 1269–1279.

Estes, S., 2005 Spontaneous mutational correlations for life-history, morphological and behavioral characters in Caenorhabditis elegans. Genetics 170: 645–653.

Estes, S., and M. Lynch, 2003 Rapid fitness recovery in mutationally degraded lines of Caenorhabditis elegans. Evolution 57: 1022–1030.

Estes, S., P. C. Phillips, and D. R. Denver, 2011 Fitness recovery and compensatory evolution in natural mutant lines of C. elegans. Evolution 65: 2335–2344.

Etienne, V., E. C. Andersen, J. M. Ponciano, D. Blanton, A. Cadavid et al., 2015 The red death meets the abdominal bristle: polygenic mutation for susceptibility to a bacterial pathogen in Caenorhabditis elegans. Evolution 69: 508–519.

Falconer, D. S., 1981 Introduction to Quantitative Genetics. Longmans Green & Co., London.

Farhadifar, R., J. M. Ponciano, E. C. Andersen, D. J. Needleman, and C. F. Baer, 2016 Mutation is a sufficient and robust predictor of genetic variation for mitotic spindle traits in Caenorhabditis elegans. Genetics 203: 1859–1870.

Félix, M.-A., and M. Barkoulas, 2015 Pervasive robustness in biological systems. Nat. Rev. Genet. 16: 483–496.

Fisher, R. A., 1930 The Genetical Theory of Natural Selection. A Complete Variorum Edition. Oxford University Press, Oxford.

Forsberg, S. K. G., J. S. Bloom, M. J. Sadhu, L. Kruglyak, and Ö. Carlborg, 2017 Accounting for genetic interactions improves modeling of individual quantitative trait phenotypes in yeast. Nat. Genet. 139: 1455.

Gaertner, B. E., M. D. Parmenter, M. V. Rockman, L. Kruglyak, and P. C. Phillips, 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.

Galton, F., 1886 Regression towards mediocrity in hereditary stature. J. Anthropol. Inst. G. B. Irel. 15: 246.

Gems, D., and D. L. Riddle, 2000   Defining wild-type life span in *Caenorhabditis elegans*. J. Gerontol. A Biol. Sci. Med. Sci. 55: B215–B219.

Ghosh, R., E. C. Andersen, J. A. Shapiro, J. P. Gerke, and L. Kruglyak, 2012   Natural variation in a chloride channel subunit confers avermectin resistance in *C. elegans*. Science 335: 574–578.

Gloria-Soria, A., and R. B. R. Azevedo, 2008   *npr-1* regulates foraging and dispersal strategies in *Caenorhabditis elegans*. Curr. Biol. 18: 1694–1699.

Goddard, M. E., K. E. Kemper, I. M. MacLeod, A. J. Chamberlain, and B. J. Hayes, 2016   Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms and genetic architecture. Proc. Biol. Sci. 283: 20160569.

Gray, J. C., and A. D. Cutter, 2014   Mainstreaming *Caenorhabditis elegans* in experimental evolution. Proc. Biol. Sci. 281: 20133055.

Greene, J. S., M. Brown, M. Dobosiewicz, I. G. Ishida, E. Z. Macosko *et al.*, 2016   Balancing selection shapes density-dependent foraging behaviour. Nature 539: 254–258.

Gruber, J. D., K. Vogel, G. Kalay, and P. J. Wittkopp, 2012   Contrasting properties of gene-specific regulatory, coding, and copy number mutations in *Saccharomyces cerevisiae*: frequency, effects, and dominance. PLoS Genet. 8: e1002497.

Gutteling, E. W., A. Doroszuk, J. A. G. Riksen, Z. Prokop, J. Reszka *et al.*, 2007   Environmental influence on the genetic correlations between life-history traits in *Caenorhabditis elegans*. Heredity 98: 206–213.

Halligan, D. L., and P. D. Keightley, 2009   Spontaneous mutation accumulation studies in evolutionary genetics. Annu. Rev. Ecol. Evol. Syst. 40: 151–172.

Hansen, T. F., 2013   Why epistasis is important for selection and adaptation. Evolution 67: 3501–3511.

Hayes, J. P., and S. H. Jenkins, 1997   Individual variation in mammals. J. Mammal. 78: 274–293.

Hemani, G., S. Knott, and C. Haley, 2013   An evolutionary perspective on epistasis and the missing heritability. PLoS Genet. 9: e1003295.

Henderson, C. R., 1985   Best linear unbiased prediction of nonadditive genetic merits in noninbred populations. J. Anim. Sci. 60: 111–117.

Hill, W. G., 1982   Rates of change in quantitative traits from fixation of new mutations. Proc. Natl. Acad. Sci. USA 79: 142–145.

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.

Hirsh, D., D. Oppenheim, and M. Klass, 1976   Development of the reproductive system of *Caenorhabditis elegans*. Dev. Biol. 49: 200–219.

Hodgkin, J., and T. Doniach, 1997   Natural variation and copulatory plug formation in *Caenorhabditis elegans*. Genetics 146: 149–164.

Houle, D., and E. J. Márquez, 2015   Linkage disequilibrium and inversion-typing of the *Drosophila melanogaster* genome reference panel. G3 (Bethesda) 5: 1695–1701.

Huang, A., S. Xu, and X. Cai, 2014   Whole-genome quantitative trait locus mapping reveals major role of epistasis on yield of rice. PLoS One 9: e87330.

Huang, B. E., A. W. George, K. L. Forrest, A. Kilian, M. J. Hayden *et al.*, 2012   A multiparent advanced generation inter-cross population for genetic analysis in wheat. Plant Biotechnol. J. 10: 826–839.

Huang, W., M. A. Carbone, M. M. Magwire, J. A. Peiffer, R. F. Lyman *et al.*, 2015   Genetic basis of transcriptome diversity in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 112: E6010–E6019.

Jiang, Y., and J. C. Reif, 2015   Modeling epistasis in genomic selection. Genetics 201: 759–768.

Johnson, N. A., 2000   Gene interactions and the origin of species, pp. 197–214 in *Epistasis and the Evolutionary Process*, edited by J. B. Wolf, E. D. Brodie, and M. J. Wade. Oxford University Press, New York.

Johnson, T. E., and W. B. Wood, 1982   Genetic analysis of lifespan in *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA 79: 6603–6607.

Jombart, T., 2008   Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24: 1403–1405.

Jombart, T., S. Devillard, and F. Balloux, 2010   Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 11: 94.

Joyner-Matos, J., A. Upadhyay, M. P. Salomon, V. Grigaltchik, and C. F. Baer, 2009   Genetic (Co)variation for life span in rhabditid nematodes: role of mutation, selection, and history. J. Gerontol. A Biol. Sci. Med. Sci. 64: 1134–1145.

Kamran-Disfani, A., and A. F. Agrawal, 2014   Selfing, adaptation and background selection in finite populations. J. Evol. Biol. 27: 1360–1371.

King, E. G., S. J. Macdonald, and A. D. Long, 2012a   Properties and power of the Drosophila synthetic population resource for the routine dissection of complex traits. Genetics 191: 935–949.

King, E. G., C. M. Merkes, C. L. McNeil, S. R. Hoofer, S. Sen *et al.*, 2012b   Genetic dissection of a model complex trait using the Drosophila synthetic population resource. Genome Res. 22: 1558–1566.

King, E. G., B. J. Sanderson, C. L. McNeil, A. D. Long, and S. J. Macdonald, 2014   Genetic dissection of the *Drosophila melanogaster* female head transcriptome reveals widespread allelic heterogeneity. PLoS Genet. 10: e1004322.

Knight, C. G., R. B. Azevedo, and A. M. Leroi, 2001   Testing life-history pleiotropy in *Caenorhabditis elegans*. Evolution 55: 1795–1804.

Kover, P. X., W. Valdar, J. Trakalo, N. Scarcelli, I. M. Ehrenreich *et al.*, 2009   A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. PLoS Genet. 5: e1000551.

Lachowiec, J., X. Shen, C. Queitsch, and Ö. Carlborg, 2015   A genome-wide association analysis reveals epistatic cancellation of additive genetic variance for root length in *Arabidopsis thaliana*. PLoS Genet. 11: e1005541.

Lessells, C. M., and P. T. Boag, 1987   Unrepeatable repeatabilities: a common mistake. Auk 104: 116–121.

Lipinski, K. J., J. C. Farslow, K. A. Fitzpatrick, M. Lynch, V. Katju *et al.*, 2011   High spontaneous rate of gene duplication in *Caenorhabditis elegans*. Curr. Biol. 21: 306–310.

Long, A. D., S. J. Macdonald, and E. G. King, 2014   Dissecting complex traits using the Drosophila synthetic population resource. Trends Genet. 30: 488–495.

Lynch, M., and B. Walsh, 1998   *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.

Macdonald, S. J., and A. D. Long, 2007   Joint estimates of quantitative trait locus effect and frequency using synthetic recombinant populations of *Drosophila melanogaster*. Genetics 176: 1261–1281.

Mackay, I. J., P. Bansept-Basler, T. Barber, A. R. Bentley, J. Cockram *et al.*, 2014   An eight-parent multiparent advanced generation inter-cross population for winter-sown wheat: creation, properties, and validation. G3 (Bethesda) 4: 1603–1610.

Malmberg, R. L., S. Held, A. Waits, and R. Mauricio, 2005   Epistasis for fitness-related quantitative traits in *Arabidopsis thaliana* grown in the field and in the greenhouse. Genetics 171: 2013–2027.

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.

Marouli, E., M. Graff, C. Medina-Gomez, K. S. Lo, A. R. Wood *et al.*, 2017   Rare and low-frequency coding variants alter human adult height. Nature 542: 186–190.

Marriage, T. N., E. G. King, A. D. Long, and S. J. Macdonald, 2014   Fine-mapping nicotine resistance loci in Drosophila using a multiparent advanced generation inter-cross population. Genetics 198: 45–57.

Masri, L., R. D. Schulte, N. Timmermeyer, S. Thanisch, L. L. Crummenerl *et al.*, 2013   Sex differences in host defence interfere with parasite-mediated selection for outcrossing during host-parasite coevolution. Ecol. Lett. 16: 461–468.

Matuszewski, S., J. Hermisson, and M. Kopp, 2015   Catch me if you can: adaptation from standing genetic variation to a moving phenotypic optimum. Genetics 200: 1255–1274.

Maupas, E., 1900   Modes et formes de reproduction des nematodes. *Arch. Zool. Exp.* 8: 463–624.

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.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al.*, 2010   The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20: 1297–1303.

McMullen, M. D., S. Kresovich, H. S. Villeda, P. Bradbury, H. Li *et al.*, 2009   Genetic properties of the maize nested association mapping population. Science 325: 737–740.

Meier, B., S. L. Cooke, J. Weiss, A. P. Bailly, L. B. Alexandrov *et al.*, 2014   *C. elegans* whole-genome sequencing reveals mutational signatures related to carcinogens and DNA repair deficiency. Genome Res. 24: 1624–1636.

Meuwissen, T., and M. Goddard, 2010   Accurate prediction of genetic values for complex traits by whole-genome resequencing. Genetics 185: 623–631.

Meuwissen, T. H., B. J. Hayes, and M. E. Goddard, 2001   Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819–1829.

Monnahan, P. J., and J. K. Kelly, 2015   Naturally segregating loci exhibit epistasis for fitness. Biol. Lett. 11: 20150498.

Morran, L. T., M. D. Parmenter, and P. C. Phillips, 2009   Mutation load and rapid adaptation favour outcrossing over self-fertilization. Nature 462: 350–352.

Morton, N. E., 1955   Sequential tests for the detection of linkage. Am. J. Hum. Genet. 7: 277–318.

Mukai, T., 1967   Synergistic interaction of spontaneous mutant polygenes controlling viability in *Drosophila melanogaster*. Genetics 61: 749–761.

Najarro, M. A., J. L. Hackett, B. R. Smith, C. A. Highfill, E. G. King *et al.*, 2015   Identifying loci contributing to natural variation in xenobiotic resistance in Drosophila. PLoS Genet. 11: e1005663.

Neher, R. A., and B. I. Shraiman, 2009   Competition between recombination and epistasis can cause a transition from allele to genotype selection. Proc. Natl. Acad. Sci. USA 106: 6866–6871.

Nigon, V., 1949   Les modalites de la reproduction et le determinisme du sexe chez quelques nematodes libres. Ann. Sci. Nat. Zool. Biol. Anim. 11: 1–132.

Noble, L. M., A. S. Chang, D. McNelis, M. Kramer, M. Yen *et al.*, 2015   Natural variation in *plep-1* causes male-male copulatory behavior in *C. elegans*. Curr. Biol. 25: 2730–2737.

Nyholt, D. R., 2000   All LODs are not created equal. Am. J. Hum. Genet. 67: 282–288.

Paaby, A. B., A. G. White, D. D. Riccardi, K. C. Gunsalus, F. Piano *et al.*, 2015   Wild worm embryogenesis harbors ubiquitous polygenic modifier variation. Elife 4: e09178.

Palopoli, M. F., M. V. Rockman, A. TinMaung, C. Ramsay, S. Curwen *et al.*, 2008   Molecular basis of the copulatory plug polymorphism in *Caenorhabditis elegans*. Nature 454: 1019–1022.

Pascual, L., N. Desplat, B. E. Huang, A. Desgroux, L. Bruguier *et al.*, 2015   Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. Plant Biotechnol. J. 13: 565–577.

Philip, V. M., G. Sokoloff, C. L. Ackert-Bicknell, M. Striz, L. Branstetter *et al.*, 2011   Genetic analysis in the Collaborative Cross breeding population. Genome Res. 21: 1223–1238.

Phillips, N., M. Salomon, A. Custer, D. Ostrow, and C. F. Baer, 2009   Spontaneous mutational and standing genetic (co)variation at dinucleotide microsatellites in *Caenorhabditis briggsae* and *Caenorhabditis elegans*. Mol. Biol. Evol. 26: 659–669.

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.

Phillips, P. C., S. P. Otto, and M. C. Whitlock, 2000   Beyond the average, pp. 20–38 in *Epistasis and the Evolutionary Process*, edited by J. B. Wolf, E. D. Brodie, and M. J. Wade. Oxford University Press, Oxford.

Poelwijk, F. J., S. Tănase-Nicola, D. J. Kiviet, and S. J. Tans, 2011   Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. J. Theor. Biol. 272: 141–144.

Poullet, N., A. Vielle, C. Gimond, S. Carvalho, H. Teotónio *et al.*, 2016   Complex heterochrony underlies the evolution of Caenorhabditis elegans hermaphrodite sex allocation. Evolution 70: 2357–2369.

Pritchard, J. K., 2002   The allelic architecture of human disease genes: common disease-common variant… or not? Hum. Mol. Genet. 11: 2417–2423.

Reddy, K. C., E. C. Andersen, L. Kruglyak, and D. H. Kim, 2009   A polymorphism in *npr-1* is a behavioral determinant of pathogen susceptibility in *C. elegans*. Science 323: 382–384.

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.

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. elegans*. Science 330: 372–376.

Rohde, P. D., B. Gaertner, K. Ward, P. Sørensen, and T. F. C. Mackay, 2017   Genomic analysis of genotype-by-social environment interaction for *Drosophila melanogaster* aggressive. Behav. Genet. 206: 1969–1984.

Ronnegard, L., X. Shen, and M. Alam, 2010   hglm: a package for fitting hierarchical generalized linear models. The R Journal 2: 20–28. Available at: http://journal.r-project.org/archive/2010-2/RJournal_2010-2_Roennegaard~et~al.pdf.

Ruby, J. G., C. Jan, C. Player, M. J. Axtell, W. Lee *et al.*, 2006   Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. Cell 127: 1193–1207.

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.

Seyfert, A. L., M. E. A. Cristescu, L. Frisse, S. Schaack, W. K. Thomas *et al.*, 2008   The rate and spectrum of microsatellite mutation in *Caenorhabditis elegans* and *Daphnia pulex*. Genetics 178: 2113–2121.

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

Simon, M., O. Loudet, S. Durand, A. Bérard, D. Brunel *et al.*, 2008   Quantitative trait loci mapping in five new large recombinant inbred line populations of *Arabidopsis thaliana* genotyped with consensus single-nucleotide polymorphism markers. Genetics 178: 2253–2264.

Sokal, R. R., and F. J. Rohlf, 1995   *Biometry: The Principles and Practice of Statistics in Biological Sciences*. W. H. Freeman, San Francisco.

Speed, D., and D. J. Balding, 2015   Relatedness in the post-genomic era: is it still useful? Nat. Rev. Genet. 16: 33–44.

Speed, D., G. Hemani, M. R. Johnson, and D. J. Balding, 2012   Improved heritability estimation from genome-wide SNPs. Am. J. Hum. Genet. 91: 1011–1021.

Speed, D., N. Cai The UCLEB ConsortiumM. Johnson, S. Nejentsev et al., 2016   Re-evaluation of SNP heritability in complex human traits. bioRxiv DOI: https://doi.org/10.1101/074310.

Sterken, M. G., L. B. Snoek, J. E. Kammenga, and E. C. Andersen, 2015   The laboratory domestication of Caenorhabditis elegans. Trends Genet. 31: 224–231.

Stewart, A. D., and P. C. Phillips, 2002   Selection and maintenance of androdioecy in Caenorhabditis elegans. Genetics 160: 975–982.

Swarup, S., W. Huang, T. F. C. Mackay, and R. R. H. Anholt, 2013   Analysis of natural variation reveals neurogenetic networks for Drosophila olfactory behavior. Proc. Natl. Acad. Sci. USA 110: 1017–1022.

Swierczek, N. A., A. C. Giles, C. H. Rankin, and R. A. Kerr, 2011   High-throughput behavioral analysis in C. elegans. Nat. Methods 8: 592–598.

Teotónio, H., S. Carvalho, D. Manoel, M. Roque, and I. M. Chelo, 2012   Evolution of outcrossing in experimental populations of Caenorhabditis elegans. PLoS One 7: e35811.

Teotónio, H., S. Estes, P. C. Phillips, and C. F. Baer, 2017   Evolution experiments with Caenorhabditis nematodes. Genetics 206: 1–27.

Theologidis, I., I. M. Chelo, C. Goy, and H. Teotónio, 2014   Reproductive assurance drives transitions to self-fertilization in experimental Caenorhabditis elegans. BMC Biol. 12: 93.

Thepot, S., G. Restoux, I. Goldringer, F. Hospital, D. Gouache et al., 2015   Efficiently tracking selection in a multiparental population: the case of earliness in wheat. Genetics 199: 609–623.

Thompson, O. A., L. B. Snoek, H. Nijveen, M. G. Sterken, R. J. M. Volkers et al., 2015   Remarkably divergent regions punctuate the genome assembly of the Caenorhabditis elegans Hawaiian strain CB4856. Genetics 200: 975–989.

Threadgill, D. W., and G. A. Churchill, 2012   Ten years of the Collaborative Cross. Genetics 190: 291–294.

Valdar, W., L. C. Solberg, D. Gauguier, S. Burnett, P. Klenerman et al., 2006   Genome-wide genetic association of complex traits in heterogeneous stock mice. Nat. Genet. 38: 879–887.

VanRaden, P. M., 2008   Efficient methods to compute genomic predictions. J. Dairy Sci. 91: 4414–4423.

Visscher, P. M., B. McEvoy, and J. Yang, 2010   From Galton to GWAS: quantitative genetics of human height. Genet. Res. (Camb.) 92: 371–379.

Visscher, P. M., G. Hemani, A. A. E. Vinkhuyzen, G.-B. Chen, S. H. Lee et al., 2014   Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. PLoS Genet. 10: e1004269.

Weigel, D., 2012   Natural variation in Arabidopsis: from molecular genetics to ecological genomics. Plant Physiol. 158: 2–22.

Weinreich, D. M., R. A. Watson, and L. Chao, 2005   Perspective: sign epistasis and genetic constraint on evolutionary trajectories. Evolution 59: 1165–1174.

Weinreich, D. M., Y. Lan, C. S. Wylie, and R. B. Heckendorn, 2013   Should evolutionary geneticists worry about higher-order epistasis? Curr. Opin. Genet. Dev. 23: 700–707.

Whitlock, M. C., and D. Bourguet, 2000   Factors affecting the genetic load in drosophila: synergistic epistasis and correlations among fitness components. Evolution 54: 1654.

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.

Wolf, J. B., E. D. Brodie, and M. J. Wade (Editors), 2000   Epistasis and the Evolutionary Process. Oxford University Press, New York.

Wood, A. R., T. Esko, J. Yang, S. Vedantam, T. H. Pers et al., 2014   Defining the role of common variation in the genomic and biological architecture of adult human height. Nat. Genet. 46: 1173–1186.

Wray, G. A., 2007   The evolutionary significance of cis-regulatory mutations. Nat. Rev. Genet. 8: 206–216.

Wright, S., 1932   The roles of mutation, inbreeding, crossbreeding, and selection in evolution. Proceedings of the Sixth International Congress of Genetics, Ithaca, New York, pp. 356–366.

Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders et al., 2010   Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42: 565–569.

Yang, J., T. A. Manolio, L. R. Pasquale, E. Boerwinkle, N. Caporaso et al., 2011   Genome partitioning of genetic variation for complex traits using common SNPs. Nat. Genet. 43: 519–525.

Young, A. I., and R. Durbin, 2014   Estimation of epistatic variance components and heritability in founder populations and crosses. Genetics 198: 1405–1416.

Zheng, C., M. P. Boer, and F. A. van Eeuwijk, 2015   Reconstruction of genome ancestry blocks in multiparental populations. Genetics 200: 1073–1087.

Zwarts, L., M. M. Magwire, M. A. Carbone, M. Versteven, L. Herteleer et al., 2011   Complex genetic architecture of Drosophila aggressive behavior. Proc. Natl. Acad. Sci. USA 108: 17070–17075.

Communicating editor: K. Nichols