

# Quantitative Epigenetics Through Epigenomic Perturbation of Isogenic Lines

Frank Johannes<sup>\*,1</sup> and Maria Colomé-Tatché<sup>†,2</sup>

<sup>\*</sup>Groningen Bioinformatics Centre, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands and <sup>†</sup>Institute for Theoretical Physics, Leibniz Universität Hannover, Appelstr. 2, D-30167, Hannover, Germany

Manuscript received January 22, 2011  
Accepted for publication February 28, 2011

## ABSTRACT

Interindividual differences in chromatin states at a locus (epialleles) can result in gene expression changes that are sometimes transmitted across generations. In this way, they can contribute to heritable phenotypic variation in natural and experimental populations independent of DNA sequence. Recent molecular evidence shows that epialleles often display high levels of transgenerational instability. This property gives rise to a dynamic dimension in phenotypic inheritance. To be able to incorporate these non-Mendelian features into quantitative genetic models, it is necessary to study the induction and the transgenerational behavior of epialleles in controlled settings. Here we outline a general experimental approach for achieving this using crosses of epigenomically perturbed isogenic lines in mammalian and plant species. We develop a theoretical description of such crosses and model the relationship between epiallelic instability, recombination, parent-of-origin effects, as well as transgressive segregation and their joint impact on phenotypic variation across generations. In the limiting case of fully stable epialleles our approach reduces to the classical theory of experimental line crosses and thus illustrates a fundamental continuity between genetic and epigenetic inheritance. We consider data from a panel of Arabidopsis epigenetic recombinant inbred lines and explore estimates of the number of quantitative trait loci for plant height that resulted from a manipulation of DNA methylation levels in one of the two isogenic founder strains.

**S**YSTEMATIC or stochastic changes in chromatin states, such as gains or losses of DNA or histone methylation, are sometimes transmitted across generations with significant phenotypic effects (RICHARDS 2006). Since different chromatin variants (epialleles) can exist in the same sequence background (*i.e.*, on the same sequence allele), they can produce a dimension of functional variation at the population level that cannot be captured by an analysis based on DNA sequence alone (JOHANNES *et al.* 2008). How much of this epigenetic variation is routinely missed in linkage or association mapping studies is an open question (JOHANNES *et al.* 2008; MAHER 2008; MANOLIO *et al.* 2009; EICHLER *et al.* 2010), but preliminary estimates in plants suggest that it can account for up to 30% of the variation in commonly studied phenotypes such as height and flowering time (JOHANNES *et al.* 2009).

Unlike DNA sequence alleles, epialleles can exhibit a high degree of instability across generations (RAKYAN *et al.* 2002; MATHIEU *et al.* 2007). Because of these dynamic properties the quantitative implications of

epigenetic inheritance in the context of human health, evolution, and agriculture have remained largely speculative (JOHANNES *et al.* 2008; RICHARDS 2008; BOSSDORF *et al.* 2008; PETRONIS 2010; BIEMONT 2010). To overcome this limitation, it is necessary to obtain a basic inventory of the transgenerational behavior of epialleles in both mammals and plants and to formally incorporate these properties into our current models of quantitative inheritance in natural and experimental populations (JOHANNES *et al.* 2008). The aim of this article is to outline both experiment and theory to achieve this.

A powerful experimental approach for studying the induction and propagation of epigenetic variation is through crosses of epigenomically perturbed isogenic strains. In the model plant Arabidopsis two groups have recently implemented such an approach by constructing so-called epigenetic recombinant inbred lines (epiRILs) (JOHANNES *et al.* 2009; REINDERS *et al.* 2009). These populations were derived from crosses between individuals with virtually identical DNA sequences but drastically divergent epigenomic profiles. In both cases, the cross was initiated from a wild-type (wt) plant and a plant carrying a single loss-of-function mutation in *ddm1* (JOHANNES *et al.* 2009) or *met1* (REINDERS *et al.* 2009), two genes involved in DNA methylation control. As a result, mutant plants exhibit significant global changes in DNA methylation (VONGS *et al.* 1993; COKUS *et al.* 2008; LISTER *et al.* 2008; REINDERS *et al.* 2009). Although mobilization of

Supporting information is available online at <http://www.genetics.org/cgi/content/full/genetics.111.127118/DC1>.

Available freely online through the author-supported open access option.

<sup>1</sup>Corresponding author: Groningen Bioinformatics Centre, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands.  
E-mail: f.johannes@rug.nl

<sup>2</sup>Present address: Groningen Bioinformatics Centre, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands.

transposable elements also occurs in these epiRILs, they nonetheless provide a unique opportunity to study the transgenerational behavior of induced epigenetic variation against a nearly invariant DNA sequence background (Figure 1A). The experimental setup for constructing such populations represents a general strategy. Similar approaches could be considered in mammals and/or through the use of environmental triggers to initiate the epigenomic changes in the parental generation (Figure 1B).

Molecular profiling of the two *Arabidopsis* epiRIL populations has shown that only a fraction of induced epialleles remain stable in subsequent generations, the rest being subject to dynamic modifications (JOHANNES *et al.* 2009; REINDERS *et al.* 2009). Two basic patterns are beginning to emerge. The first pattern indicates that a subset of epialleles undergo rapid and stochastic fluctuations over a wide spectrum of chromatin states, many of which are outside of the parental range (REINDERS *et al.* 2009). While such alterations can be causative of phenotypes within a given generation, they probably do not contribute to phenotypic inheritance (SLATKIN 2009) and can therefore be regarded as noise in the underlying heritable substrate. The second, and more important pattern, is a systematic and gradual reversion of mutant epiallelic states to those of the wt over the course of several generations. This process represents an intrinsic rescue system that is invoked to restore proper genome function and integrity. Loci that meet this pattern tend to correspond to sequences that are continuously targeted by the RNA-directed DNA methylation (RdDM) machinery (JOHANNES *et al.* 2009; TEIXEIRA *et al.* 2009; TEIXEIRA and COLOT 2010).

Epiallelic instabilities, as described above, create a complex source of heritable variation. These properties pose challenges to the way we have to approach quantitative inheritance in the epiRIL or similar populations. The key task is to simultaneously account for two processes: The first involves the meiotic transmission of maternal and paternal DNA sequence haplotypes according to Mendelian laws. The second is a dynamic process that governs continuous changes in the chromatin states (epialleles) harbored by these haplotypes and leads to non-Mendelian patterns of inheritance. Here we develop the necessary theoretical foundation to quantify these processes using epiRILs (Figure 1A) as a model system. We find that epiallelic reversion, recombination, parent-of-origin effects, and transgressive segregation are key parameters in these populations: Their joint effects can produce complex and highly dynamic inheritance patterns that cannot be predicted from strictly Mendelian models. In the limiting case of fully stable epialleles our model reduces to the classical theory of experimental line crosses and thus illustrates a fundamental continuity between genetic and epigenetic inheritance. In what follows we present the first comprehensive attempt to quantify epigenetic inheritance in model organisms.

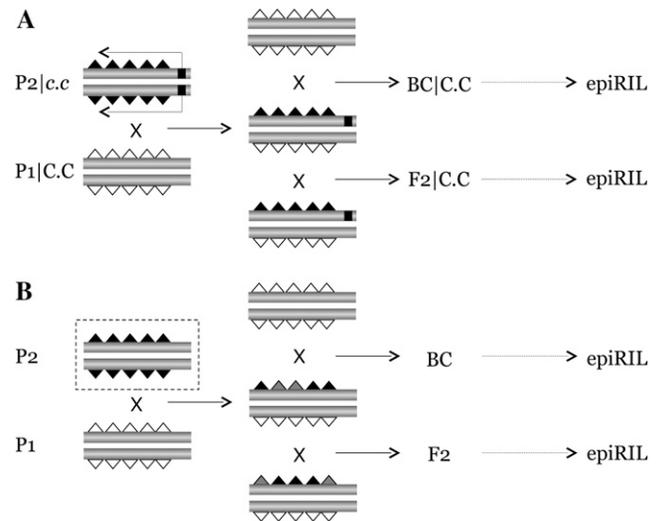


FIGURE 1.—Construction of epigenetic recombinant inbred lines (epiRIL): (A) Induction of epigenomic perturbation by means of a mutation in genes involved in chromatin control, followed by selfing (plants) or sibling mating (mammals) of conditional intercross ( $F_2|C.C$ ) or backcross ( $BC|C.C$ ) progeny. It is assumed that the parental strains have nearly identical DNA sequences (shaded horizontal bars) but different epigenomic profiles (white, gray, and black triangles). The exceptions are rare *de novo* sequence changes (not shown for simplicity) resulting from compromised chromatin states in the mutant parent ( $P_2|c.c$ ). (B) Same crossing scheme as above but using environmental manipulations in place of mutations to invoke the initial perturbation.

## THEORY

**Conceptual basis:** Consider a locus,  $L$ , extensively involved in genome-wide chromatin control. Genotype C.C at this locus corresponds to proper chromatin maintenance, whereas the mutant genotype  $c.c$  induces global chromatin changes (*e.g.*, modifications of DNA or histone methylation). We start with two inbred parents, a wild-type parent,  $P_1|C.C$ , and a mutant parent,  $P_2|c.c$ . By design, the two parents have identical DNA sequences (except at locus  $L$  and inevitably at a small number of other loci; MIROUZE *et al.* 2009; TSUKAHARA *et al.* 2009), but drastically divergent chromatin profiles (Figure 1A).

As a result of the epigenomic perturbation induced by the mutation, the two parents will differ in their chromatin states at  $N$  loci determining a quantitative trait  $y$ . Suppose that in the  $P_1|C.C$  parent  $N(1 - \tau)$  of these loci have stable epigenotype  $\Omega.\Omega$  and  $N\tau$  of the loci have stable epigenotype  $\omega.\omega$  (Figure 2A). Here epiallele  $\Omega$  corresponds to a phenotypically increasing and  $\omega$  to a decreasing chromatin state. Relative to  $P_1|C.C$ , the mutant parent,  $P_2|c.c$ , will have undergone the following possible epiallelic changes at the  $N$  loci:  $\Omega \rightarrow \omega$ ,  $\Omega \rightarrow \tilde{\omega}$ ,  $\omega \rightarrow \Omega$ , and  $\omega \rightarrow \tilde{\Omega}$ , where the tilde ( $\tilde{\phantom{x}}$ ) signifies an unstable epiallelic state (Figure 2A).

We suppose that a proportion  $s$  of the newly induced epialleles remain stable in subsequent generations ( $\Omega$

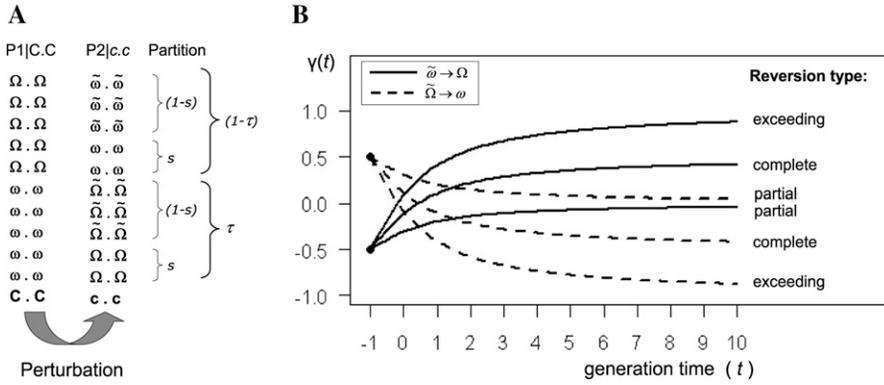
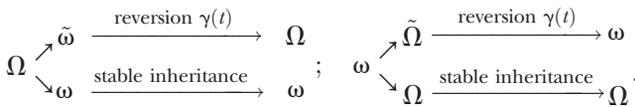


FIGURE 2.—Epigenomic structure of the parental strains and epiallelic reversion: (A) As a result of the perturbation the wt ( $P1|C.C$ ) and the mutant ( $P2|c.c$ ) parents will differ in their diploid chromatin states (epigenotypes) at  $N$  loci. The mutant ( $P2|c.c$ ) parent will have a proportion  $\tau$  of phenotypically increasing epigenotypes ( $\Omega, \tilde{\Omega}$  and  $\tilde{\Omega}, \tilde{\Omega}$ ) as well as a proportion  $(1 - \tau)$  of decreasing epigenotypes ( $\omega, \tilde{\omega}$  and  $\tilde{\omega}, \tilde{\omega}$ ), with  $s$  and  $(1 - s)$  of these being either stable or unstable, respectively (see text). (B) Unstable epialleles ( $\tilde{\Omega}, \tilde{\omega}$ ) induced in the mutant parent have the capacity to revert to wt states

over time according to some function,  $\gamma(t)$ . This reversion can be perfect or imperfect. We quantify epialleles in the parental generation by setting  $\Omega = \tilde{\Omega} = \frac{1}{2}$  and  $\omega = \tilde{\omega} = -\frac{1}{2}$ . Starting at  $t = 0$  (the  $F2|C.C$  or  $BC|C.C$  generation) the state values of reversible epialleles begin to change.

and  $\omega$ ) (Figure 2A). This is even the case when proper chromatin maintenance function is restored. Since stable epialleles behave like DNA sequence changes at the population level, we make no formal distinction between them. Only direct molecular profiling and sequencing of the two parents and their cross-derivatives will make it possible to uncover the physical basis of such stable induced alterations.

Apart from stable epialleles, we also assume a proportion  $(1 - s)$  of newly induced epialleles ( $\tilde{\Omega}$  and  $\tilde{\omega}$ ) with the capacity to revert to approximate wild-type states over generations. We quantify this physical process through a function  $\gamma(t)$ , which describes the progressive changes of epiallelic states in continuous time (Figure 2B). Specifically,  $\tilde{\Omega}$  and  $\tilde{\omega}$  correspond to sequences that are targeted by the RdDM machinery (JOHANNES *et al.* 2009; REINDERS *et al.* 2009; TEIXEIRA *et al.* 2009; TEIXERA and COLOT 2010) or possibly other correction mechanism. For the cross design shown in Figure 1A, this reversion begins after the C.C genotype has been reintroduced at  $L$ , that is, in the conditional F2 ( $F2|C.C$ ) or backcross ( $BC|C.C$ ) populations (Figure 2, A and B). Upon further propagation of individual lines from these conditional populations by means of selfing (plants) or sibling mating (mammals), progressive epiallelic reversion continues through recurrent maintenance action at each generation. Including the initial perturbation, the different epiallelic fates outlined above can be summarized schematically as follows:



Our goal is to model this process for any generation of inbreeding. This allows us to draw direct connections between the basic properties of epialleles and their impact on heritable variation at the population level.

**Transgenerational epigenetic dynamics:** We quantify the epigenotype at locus  $j$  at time  $t$  using the coding introduced in Table 1.

*Parental generation:* Since all individuals are assumed homozygous in the parental generation, the total phenotypic variance can be expressed as

$$\sigma_P^2(y) = \sigma_\epsilon^2 + [N\delta(2\tau - 1)]^2, \quad (1)$$

where  $\sigma_\epsilon^2$  is the pooled within-line (environmental) variance, and  $\delta$  is henceforth defined as the average contribution of a single locus to the between-parental phenotypic mean difference  $D$  (SEREBROVSKY 1928):

$$\delta = \frac{D}{2N(2\tau - 1)}. \quad (2)$$

This expression assumes that  $\delta$  is equivalent over all  $N$  causative loci; that is,  $\delta_1 = \delta_2 = \dots = \delta_N = \delta$ .

Note in Equation 1 that when the transgression parameter  $\tau = 0.5$  we have that  $\sigma_P^2(y) = \sigma_\epsilon^2$ . In this limiting case the total phenotypic variance in the parental generation is purely environmental, despite drastic functional divergence between the parents. This situation provides the condition for a maximum gain in transgressive variance in subsequent generations. It follows from the fact that each parent is fixed for both phenotypically decreasing and increasing states so that recombination events can produce offspring with more extreme epigenotypes (RIESENBERG *et al.* 1999). We therefore refer to parameter  $\tau$  as a measure of “transgression potential” in the parental generation, which can become “realized,” in some sense, in subsequent generations. This phenomenon is discussed in detail below.

*F1 and base-population:* Crossing  $P1|C.C \times P2|c.c$  yields the  $F1$  generation. We assume that epiallelic states

**TABLE 1**  
**Coding of epigenotypes at a locus**

Epigenotype	Value		
	Additivity	Mutant dominance	wt dominance
$\Omega.\Omega$	+1	+1	+1
$\omega.\omega$	-1	-1	-1
$\Omega.\omega$	0	+1 (if $\Omega$ is the mutant) -1 (if $\omega$ is the mutant)	-1 (if $\Omega$ is the mutant) +1 (if $\omega$ is the mutant)
$\tilde{\Omega}.\tilde{\Omega}$	$2\gamma(t)$	$2\gamma(t)$	$2\gamma(t)$
$\tilde{\omega}.\tilde{\omega}$	$-2\gamma(t)$	$-2\gamma(t)$	$-2\gamma(t)$
$\tilde{\Omega}.\tilde{\omega}$	$1/2 + \gamma(t)$	$2\gamma(t)$	-1
$\tilde{\omega}.\tilde{\Omega}$	$1/2 - \gamma(t)$	$-2\gamma(t)$	+1

Shown is the coding used for the different epigenotypes at a given locus at time  $t$ . We distinguish between additivity, complete dominance due to mutant parental epialleles, and complete dominance due to wt parental epialleles.

induced in the  $P2|c.c$  parent remain stable (TEIXEIRA *et al.* 2009), but this assumption can be easily relaxed if necessary. As a consequence, the phenotypic variance may or may not (RICHARDS 2009) be equivalent to the environmental variance. As shown in Figure 1A the  $F1|C.c$  are used to derive the base population for advanced inbreeding generations either through a backcross ( $F1|C.c \times P1|C.C$ ) or through an intercross ( $F1|C.c \times F1|C.c$ ). From these crosses, only the C.C progeny ( $BC|C.C$  or  $F2|C.C$ ) are selected to initiate the inbreeding process through selfing or sibling mating. This permits a detailed study of the time-dependent behavior of parental epialleles independent of the recurrent action of the  $c.c$  genotype. For simplicity we ignore the introgression of wt epigenotypes surrounding locus  $L$  as a result of this selection procedure.

*Advanced inbreeding generations:* At any time point  $t$  of inbreeding, the variance in trait  $y$  is the sum of an epigenetic component,  $\sigma^2(\eta, t)$ , and an environmental (error) component,  $\sigma^2(\varepsilon)$ :

$$\sigma^2(y, t) = \sigma^2(\eta, t) + \sigma^2(\varepsilon). \quad (3)$$

We assume that epigenotypes are uncorrelated with the error term  $\varepsilon$ , and that the error terms are uncorrelated across generation times. With equal effect sizes across the  $N$  causative loci, the epigenetic variance can be approximated as

$$\sigma^2(\eta, t) \simeq \delta^2 N \{ \sigma^2(\eta_j | t) + (N-1) \sigma(\eta_j, \eta_k | t, \bar{r}) \}, \quad (4)$$

where  $\sigma^2(\eta_j | t)$  is the variance at a single locus  $j$  at time  $t$  and  $\sigma(\eta_j, \eta_k | t, \bar{r})$  is the covariance between any two loci  $j$  and  $k$  separated by an average pairwise recombination fraction  $\bar{r}$  (FRANKLIN 1970). It can be shown (see Appendix A) that Equation 4 has the explicit form

$$\begin{aligned} \sigma^2(\eta, t) \simeq \frac{N\delta^2}{4} \{ & q_1 [4\gamma(\gamma+1)(s-1) - 3s - 1] \\ & + (N-1)q_2 [1+s+2\gamma(1-s)]^2 (1-2\tau)^2 \}, \end{aligned} \quad (5)$$

where we put  $\gamma = \gamma(t)$  to lighten the notation. The parameters  $q_1$  and  $q_2$  depend on the mode of action (additivity or dominance), the type of inbreeding scheme (selfing or sibling mating), and the base population (backcross or F2-intercross) used to initiate the inbreeding processes. Table 2 provides a summary of the specific form of  $q_1$  and  $q_2$  in each of these cases.

An important observation is that dominance (complete dominance in our case) appears as a time-dependent phenomenon, owing not only to the progressive depletion of heterozygote epigenotypes, but also to the reversion of mutant epialleles to wt states. It is therefore necessary to distinguish two types of dominance effects, one being attributable to epialleles inherited from the mutant parent ( $P2|c.c$ ) and the other being due to epialleles deriving from the wt parent ( $P2|C.C$ ) (see Tables 1 and 2). As we show, this distinction has an effect on the epigenetic variation when inbreeding is carried forward from a backcross base population, as a result of the initial asymmetry of the epigenotype frequencies. In the case of a F2 base population, on the other hand, the epigenetic variance component is equivalent under the two dominance scenarios (what differs is the phenotypic mean of the population).

**Estimation of the number of induced quantitative trait loci (QTL):** As an extension of previous biometrical approaches (CASTLE 1921; SEREBROVSKY 1928; LANDE 1981; ZENG 1992), Equation 5 can be used directly to obtain conservative estimates of the number of QTL ( $N$ ) resulting from the initial epigenomic perturbation in the parental generation. To achieve this, we substitute  $\delta$  from Equation 2 into the equation for the

**TABLE 2**  
Parameter values  $q_1$  and  $q_2$

Base	Effect	Parameter values
Selfing		
F2	Add.	$q_1 = \frac{1}{2^{t+1}} - 1$ $q_2 = \frac{2\bar{r} - 1}{2\bar{r} + 1} \left( \frac{(1 - 2\bar{r})^{t+1}}{2^{t+1}} - 1 \right)$
F2	Dom.	$q_1 = \frac{1 - 2^{2+2t}}{2^{2t+2}}$ $q_2 = \frac{1}{2^{t+1}} \left( (1 + 2\bar{r}(\bar{r} - 1))^t \frac{(1 - 2\bar{r}^2 + 4\bar{r}^3)}{1 + 2\bar{r}} - \frac{(1 - 2\bar{r})^{t+2}}{1 + 2\bar{r}} \right) + \frac{1 - 2\bar{r}}{1 + 2\bar{r}} - \frac{1}{2^{t+2}}$
BC	Add.	$q_1 = \frac{1}{2^{t+1}} - \frac{3}{4}$ $q_2 = \frac{2\bar{r} - 1}{2\bar{r} + 1} \left( \frac{(1 - 2\bar{r})^t (1 - \bar{r})}{2^{t+1}} - \frac{3}{4} \right)$
BC	Dom.	$q_1 = \frac{1 + 2^{t+1}\xi - 3 \times 2^{2t}}{2^{2t+2}}, \xi = 1 \rightarrow \text{wt dominant}, \xi = -1 \rightarrow \text{mutant dominant}$ $q_2 = -\frac{1}{2^{t+1}} \left( \frac{(1 - 4\bar{r}^2)\xi}{2\bar{r} + 1} + (\bar{r} - 1) \left[ (1 + 2\bar{r}(\bar{r} - 1))^t - \frac{(1 - 2\bar{r})^{t+1}}{1 + 2\bar{r}} \right] \right)$ $- \frac{3}{4} \frac{2\bar{r} - 1}{2\bar{r} + 1} - \frac{1}{2^{t+2}},$ $\xi = 1 \rightarrow \text{wt dominant}, \xi = -1 \rightarrow \text{mutant dominant}$
Sib.mat		
F2	Add.	$q_1 = -1 - \frac{1}{4} \left( \frac{1 - \sqrt{5}}{4} \right)^t \left( \frac{3}{\sqrt{5}} - 1 \right) + \frac{1}{4} \left( \frac{1 + \sqrt{5}}{4} \right)^t \left( \frac{3}{\sqrt{5}} + 1 \right)$
F2	Dom.	$q_1 = \frac{1}{5} \left( \frac{(-1)^{t+1}}{2^{1+2t}} - \frac{1}{2^{3+4t}} \left[ (3\sqrt{5} - 7)(1 - \sqrt{5})^{2t} - (3\sqrt{5} + 7)(1 + \sqrt{5})^{2t} \right] - 5 \right)$
BC	Add.	$q_1 = -\frac{3}{4} - \frac{1}{4} \left( \frac{1 - \sqrt{5}}{4} \right)^t \left( \frac{2}{\sqrt{5}} - 1 \right) + \frac{1}{4} \left( \frac{1 + \sqrt{5}}{4} \right)^t \left( \frac{2}{\sqrt{5}} + 1 \right)$
BC	Dom.	$q_1 = \frac{1}{5} \left( \frac{(-1)^t}{2^{3+2t}} + \frac{1}{2^{2+2t}} \left[ \xi(5 - 2\sqrt{5})(1 - \sqrt{5})^t + \xi(5 + 2\sqrt{5})(1 + \sqrt{5})^t \right. \right.$ $\left. \left. + \frac{1}{2^{2+2t}} ((9 - 4\sqrt{5})(1 - \sqrt{5})^{2t} + (9 + 4\sqrt{5})(1 + \sqrt{5})^{2t}) \right] - \frac{15}{4} \right),$ $\xi = 1 \rightarrow \text{wt dominant}, \xi = -1 \rightarrow \text{mutant dominant}$

The analytical form for the parameters  $q_1$  and  $q_2$  in Equation 5 is specified for the different combinations of mating scheme (selfing or sibling mating), base-population (Base), additivity (Add.), or complete dominance (Dom.) effects (see Appendix A). For sibling mating,  $q_2$  can be exactly evaluated numerically for any given value of  $\bar{r}$ .

epigenetic variance (Equation 5), and solve for  $N$  at any generation  $t$  of inbreeding to obtain

$$N = \frac{D^2 \left\{ q_1 [1 - 4\gamma(1 + \gamma)(s - 1) + 3s] + q_2 [1 - 2\gamma(s - 1) + s]^2 (1 - 2\tau)^2 \right\}}{(1 - 2\tau)^2 \left\{ D^2 q_2 [1 - 2\gamma(s - 1) + s]^2 - 16\sigma^2(\eta, t) \right\}}. \quad (6)$$

It is perhaps interesting to note that if we consider the restrictive case of a F2 base population with

additivity,  $s = 1$  (fully stable epialleles),  $\tau = 0$  (no transgression),  $\bar{r} = 0.5$  (linkage equilibrium among  $N$  loci), and  $t = 0$ , Equation 6 reduces to the well-known Castle-Wright estimator (CASTLE 1921).

The most important result of this manuscript is Equation 5. It formalizes the relationship between epiallelic reversion (via  $\gamma(t)$  and  $s$ ), recombination (via  $\bar{r}$ ),

transgression (via  $\tau$ ) and parent-of-origin effects (by keeping track of epiallelic origins). This relationship jointly determines the epigenetic variation in the population during inbreeding. In the following section we examine more closely the complex and highly dynamic patterns of heritable variation that can arise from it.

## RESULTS

Classical theory of experimental line crosses typically assumes no transgression ( $\tau = 0$ ) and the transmission of fully stable parental alleles ( $s = 1$ ). Sometimes, it is further assumed that all loci are in linkage equilibrium ( $\bar{r} = 0.5$ ). These constraints represent special cases of the theory developed here. We treat these scenarios as a reference against which to compare the rich spectrum of epigenetic inheritance in epigenomically perturbed line crosses. The phenotypic variance at any time point is simultaneously determined by all of the parameters specified in Equation 5. For clarity we assess their influences systematically by varying them one at the time. Several key population-level phenomena are considered for the case of selfing starting from a backcross base population. The case of selfing from a F2 base population can be found in the [supporting information](#). Throughout, we fix the average recombination rate ( $\bar{r}$ ) at 0.44, a value that is based on the Arabidopsis genetic map (LYNCH and WALSH 1998).

**Inheritance of unstable alleles:** We first consider the case of complete epiallelic instability ( $s = 0$ ) and no transgression ( $\tau = 0$ ). In this case all induced mutant epialleles are effectively reverted to the wt state over time. The specific form of the reversion function that governs this process is currently unknown, but should be of substantial interest in future empirical studies (JOHANNES *et al.* 2008). Although it is reasonable to assume that reversion is locus specific, for the purpose of providing an average description of the system it suffices to consider an average (between-locus) reversion function,  $\gamma(t)$ . We tentatively posit the form  $\gamma(t) = 1/2 - 2/\pi \arctan(\beta(t+1))$ , where  $\beta$  is the rate parameter (Figure 2B). We vary  $\beta$  so that we can examine the full range from slow to fast reversion. In contrast to stable Mendelian inheritance (Figure 3, A–C (I), black solid and dashed line), the reversion function has the effect of eroding heritable variation over time so that as  $t \rightarrow \infty$  the heritable variation in the population is progressively lost (Figure 3, A–C (I), dark gray solid lines).

**Mixed inheritance of stable and unstable epialleles:** Complete instability of epialleles is an extreme case. It is more likely that a proportion of the induced epialleles remain stable and produce Mendelian inheritance patterns (JOHANNES *et al.* 2009; REINDERS *et al.* 2009). Indeed, preliminary molecular profiling of the *ddm1*-derived epiRILs suggests that only up to one-half of the tested epialleles are reversible with the remaining loci being stable for at least eight generations of inbreeding

(JOHANNES *et al.* 2009; TEIXEIRA *et al.* 2009). Moreover, gross perturbations of the epigenome can lead to *de novo* sequence variation through the insertion of remobilized transposable elements or other structural abnormalities. These induced alterations in the parental generation also contribute to the fraction of stable segregating variation. The precise proportions depend on the particular perturbation, organism, and experimental setup and needs to be determined on a case-by-case basis.

We explore the effect of different proportions assuming a fixed reversion function and no transgression,  $\tau = 0$ . Figure 3, A–C (II), illustrates the effect of this type of mixed inheritance with the phenotypic variance being decreased over generation time due to reversion but converging to a stable value as  $t \rightarrow \infty$ . The final variance represents the stable heritable substrate that can be gleaned from the initial perturbation. It should be clear that as the proportion of stable epialleles in the genome approaches unity ( $s \rightarrow 1$ ), our model converges to the familiar case of Mendelian inheritance, which is the basis of the classical theory of experimental line crosses.

**Effects of imperfect epigenomic resetting:** The reversion of mutant epialleles to the wt state can be an imperfect process (Figure 2B). It is possible that epialleles converge to values outside of the parental range (REINDERS *et al.* 2009). For example, the remethylation of *hypomethylated* mutant epialleles could produce *hypermethylated* states in subsequent generations relative to the wt (REINDERS *et al.* 2009). We consider this case by letting all unstable epialleles revert to state values that are twice that of the wt parent. This implies that the reversion of mutant states must first pass through wt states before they reach their final stable state. This process leaves an interesting signature at the population level: It leads to an initial depletion of epigenetic variance, before there is an unexpected gain in heritable variance at later generations (Figure 3, A–C (III) light gray solid lines).

A different pattern occurs in situations where epiallelic reversion converges to intermediate parental values (0.5 of wt values). In this case, heritable variation is never completely lost (Figure 3, A–C (III), dark gray solid lines). Note that this latter pattern mimics the situation observed for the case of mixed inheritance (previous section). Distinguishing these two possibilities empirically therefore depends on detailed knowledge of the average reversion function,  $\gamma(t)$ , and the proportion of stable epialleles,  $s$ , operating in a given population.

**Realized transgression potentials:** It has been shown that transgressive segregation is widespread in experimental line crosses (RIESENBERG *et al.* 1999). A major reason for this is that the parents have often not undergone divergent selection prior to crossing and are therefore each fixed for both increasing and decreasing genotypes. In QTL experiments this is reflected in the detection of QTL with opposite signs relative to the parental means.

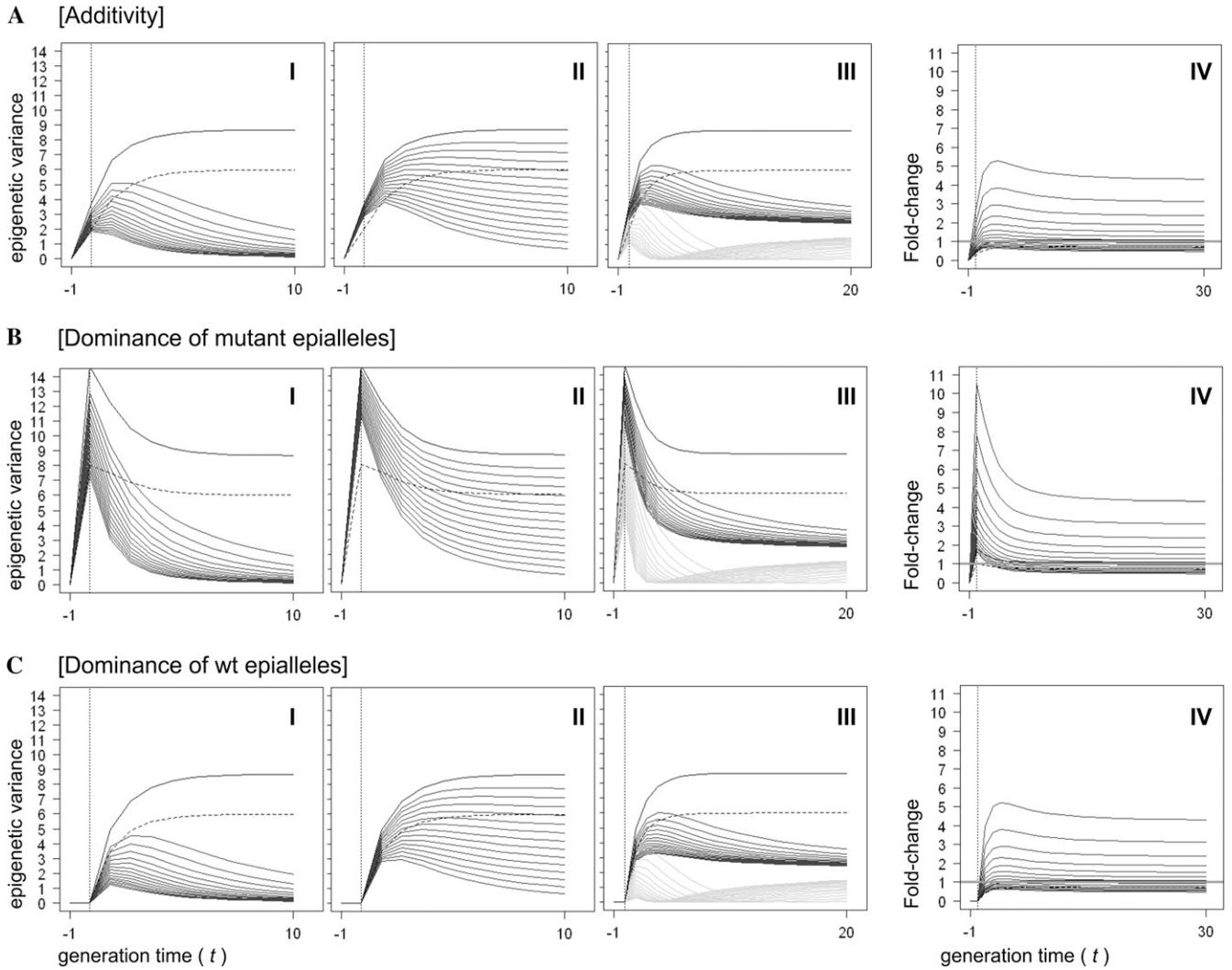


FIGURE 3.—Transgenerational dynamics of epigenetic variation. We show the dynamic behavior of epigenetic variation in the case of additivity (A), complete dominance of mutant epialleles (B), and complete dominance of wt epialleles (C). Throughout we show the Mendelian inheritance case with linkage ( $\bar{r} = 0.44$ ,  $s = 1$  and  $\tau = 0$ , solid black line) and with no linkage ( $\bar{r} = 0.5$ ,  $s = 1$  and  $\tau = 0$ , dashed black line). For each mode of action we consider four key phenomena: inheritance of unstable epialleles (A–C, I):  $s = 0$ ,  $\tau = 0$ , and variable reversion function with rate parameter  $\beta$  evaluated at equal increments over the range 0.1 (top line) to 0.5 (bottom line);  $x$ -axis plots generation time and  $y$ -axis gives the epigenetic variance,  $\sigma^2(\eta, t)$ . Mixed inheritance of stable and unstable epialleles (A–C, II):  $\tau = 0$ , fixed reversion function with  $\beta = 0.2$ , and variable  $s$  evaluated at equal increments over the range 0 (bottom line) to 0.9 (top line);  $x$ -axis and  $y$ -axis defined as above. Effects of imperfect epigenomic resetting (A–C, III): Parameter settings were chosen as in (A–C, I) but reversion was assumed to be imperfect with transitions to epiallelic states that are twice the initial wt state (dark gray solid lines) or intermediate between wt and mutant (light gray solid lines);  $x$ -axis and  $y$ -axis defined as above. Realized transgression potentials (A–C, IV):  $s = 0.5$ , fixed reversion function with  $\beta = 0.2$ , and variable  $\tau$  evaluated in equal increments over the range 0 (bottom line) to 0.35 (top line); the  $x$ -axis plots generation time and the  $y$ -axis gives the fold change in epigenetic variance relative to the between-parental variance,  $\sigma^2(\eta, t)[\sigma_p^2(y) - \sigma_e^2]^{-1}$ .

In perturbation-derived populations as the ones described here, transgressive segregation is likely to be a key aspect of quantitative inheritance. The removal of methylation, for instance, can lead to active or inactive chromatin states in the mutant parent, which can translate into increasing or decreasing phenotypic values, depending on whether the underlying sequences are involved in inhibitory or facilitating functions in the networks that connect (epi)genotype with phenotype (Figure 2A).

We explore the effect of different transgression potentials ( $\tau$ ) for a fixed reversion function and  $s = 0.5$ . We vary  $\tau$

between 0 and 0.35 ( $\tau = 0.5$  being the maximum). The effect for large  $\tau$  is dramatic. Under additivity, and assuming  $\tau = 0.35$ , we find that heritable variation increases in the order of fivefold relative to the between parental variance (Figure 3A, IV). This effect is further exaggerated when mutant epialleles act dominantly. In this case we observe a nearly 11-fold increase in heritable variation at early generations (Figure 3B, IV) followed by a gradual decrease due to the progressive loss of heterozygote epigenotypes.

Our treatment of transgression represents the first theoretical approach to highlight the importance of

transgressive segregation in generating phenotypic innovation in experimental line crosses.

**Application:** A recent analysis of the *ddm1*-derived Arabidopsis epiRILs reported large heritable variation for plant height. This work has shown that perturbations of plant methylomes are sufficient to induce lasting phenotypic consequences in commonly studied complex traits. An important question concerns the specific features of the heritable architecture that has been set up in the epiRILs, such as its physical basis (*e.g.*, sequence *vs.* methylation based) or the number and sizes of induced QTL. Definitive answers to this question require genome-wide epigenetic profiling techniques (*e.g.*, ChIP-chip, ChIP-seq, or BS-seq) in combination with QTL mapping methods as well as resequencing of each line (JOHANNES *et al.* 2009). While such efforts are underway, we here consider using Equation 6 as the basis for deriving the first estimates of the number of QTL (and their average effects sizes) underlying plant height in this population.

Since phenotypic data from only two time points is available (parental generation and generation eight of inbreeding), it is necessary to make informed assumptions about several of the unknown parameters specified in Equation 6. These assumptions rely heavily on previous molecular or phenotypic observations of this population (JOHANNES *et al.* 2009; TEIXEIRA *et al.* 2009) and are explicitly listed below:

- $\sigma^2(\eta, t) = 11.2$ : estimated from a random-effects model.
- $D = 9.92$ : difference in phenotypic means (in centimeters) between the *ddm1* and wt parents.
- $s = 0.5$ : based on an analysis of a random subset of loci.
- $\bar{r} = 0.44$ : average recombination fraction for the Arabidopsis genome (LYNCH and WALSH 1998).

Values for the transgression potential,  $\tau$ , remain elusive because this quantity cannot be directly measured using molecular techniques. Keeping this limitation in mind, we show various estimates for the number of QTL ( $N$ ) for different values of  $\tau$  (Figure 4). They range from as low as 1 (for  $\tau = 0$ ) to as high as 6 (for  $\tau = 0.31$ ). A unique estimate of  $N$  can be obtained by fixing  $\tau$  at its theoretical average (see Appendix B). In this case we find that  $\bar{N} = 2.26$  (95% CI =  $\pm 0.31$ ) (Figure 4, black bar). This (conservative) estimate suggests the induction of a polygenic heritable architecture, with each QTL explaining  $\sim 14\%$  of the phenotypic variance in plant height. Given these considerably large effect sizes, the underlying causative loci should be mappable in future integrative QTL studies, even with relatively small sample sizes.

## DISCUSSION

Epigenetic modifications, such as DNA methylation, are not only widely conserved across species (ZEMACH

*et al.* 2010; FENG *et al.* 2010), but also show substantial interindividual variation within populations (VAUGHN *et al.* 2007; ZHANG *et al.* 2008; KAMINSKY *et al.* 2009). The possibility that this type of epigenetic variation can influence phenotypes independent of DNA sequence variation poses major challenges to our current understanding of complex trait inheritance.

The first problem is that sequence-based mapping approaches (*i.e.*, linkage or association mapping) may be insufficient to fully capture the heritable architecture of complex traits (JOHANNES *et al.* 2008). A recent study by KONG *et al.* (2009), for example, illustrates that knowledge of the epigenetic status of sequence alleles (in this case, the parent-of-origin of the alleles) is often necessary to establish significant associations with phenotypes. Although the authors found that these effects were sex specific, it raises the need to routinely include both levels of variation (DNA sequence or chromatin state) in the analysis.

The second and partially related problem is that the potential temporal instability of epigenetic variation can produce a level of phenotypic dynamics at the population-level that cannot be predicted from strictly Mendelian models of inheritance. Here we have outlined an experimental and theoretical approach to begin to address this issue. The approach relies on the use of a perturbation strategy to induce epigenetic variation in isogenic lines, followed by a transgenerational assessment of derivative populations. This provides an ideal platform for studying the temporal properties of epialleles and permits a theoretical description of these properties in connection with the inheritance of complex traits observed in such populations.

**Extensions using environmental triggers:** Two experimental examples of this approach have recently been implemented in the model plant Arabidopsis using loss-of-function mutations in *ddm1* and *met1*, two genes involved in genome-wide methylation maintenance, to initiate the perturbation. While this continues to be a valuable resource, future experiments should attempt to produce similar populations using environmental manipulations in place of mutations (Figure 1B). BOSSDORF *et al.* (2010), for instance, demonstrated that treatment of Arabidopsis ecotypes with demethylation agents is sufficient to invoke phenotypically relevant methylation changes. Similarly, VERHOEVEN *et al.* (2010) used environmental stressors, such as chemical induction of herbivore and pathogen defenses, and noted heritable DNA methylation changes in asexual dandelion. These types of interventions may be sufficient to induce and propagate epigenetic variation in the parents and their cross derivatives. A demonstration of this should have important implications for evolutionary theory, which traditionally draws a clear divider between the environment and the heritable material. There is certainly need for modeling approaches that incorporate environmentally induced epigenetic changes into evolutionary

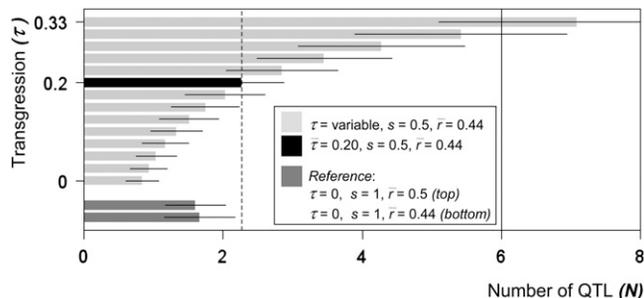


FIGURE 4.—Estimates of the number of QTL. Plotted are estimates of number of QTL ( $N$ ) against the transgression potential parameter  $\tau$ . Parameters were set according to the shade coding shown in the figure. Confidence intervals (95%) were calculated using 1000 nonparametric bootstrap samples (Appendix B). The black solid vertical line marks the expected maximum detectable  $N$  in a backcross-derived RIL, which is approximately 6 (see Appendix B).

theory. More general attempts to assess the role of epigenetic inheritance in the context of selection and adaptation have been undertaken (CSABA 1998; PÁL and MIKLÓS 1999; BONDURIANSKY and DAY 2009). The inclusion of epiallelic reversion, as we have formalized it here, would be an appealing extension. It is tempting to speculate that such reversion processes have evolved to facilitate short-term adaptation of populations to rapid environmental changes. These issues are outside of the scope of this work.

**Considerations for mammalian populations:** Studies of epigenetic inheritance have been primarily pursued in plant species, and much less is known about the transgenerational behavior of epialleles in mammalian populations. The dominant paradigm dictates that the epigenome is completely reset during early mammalian development (REIK 2007; FENG *et al.* 2010), which implies that induced epigenetic effects are not carried to subsequent generations. Formally, this suggests a reversion process that follows a step function, dropping to wt levels at  $t = -1$  (first-generation progeny of two parents). However, several well-documented single-locus examples of epiallelic inheritance exist in mammalian systems (RAKYAN *et al.* 2003; BLEWITT *et al.* 2006; DAXINGER and WHITELAW 2010). These findings are probably no exception as more recent genome-wide surveys of mouse gametes show clear instances of transmitted parental DNA methylation profiles (BORGEL *et al.* 2010), suggesting that epigenetic inheritance may be much more widespread in mammalian populations than previously acknowledged.

By help of transgenerational phenotypic data, concrete hypotheses about the extent of epigenomic resetting can be tested on statistical grounds using the experimental framework outlined in this article. This can be achieved by considering alternative reversion functions in a fit to the data. Such a proposal is akin to the approach recently developed by TAL *et al.* (2010), which permits hypothesis tests about effect of epi-

genomic resetting events on the covariance between relatives.

We argue that the construction of mammalian crosses between isogenic parents with perturbed and unperturbed epigenomes will be critically important to begin to extrapolate results to humans. The theory outlined in this article, particularly the results for sibling mating, is entirely compatible with an analysis of experimental mammalian populations (*e.g.*, mice or rats). However, the initial construction of such crosses using perturbations may be more complicated than in plants, given that mutations in genes controlling DNA methylation tend to be lethal (LI *et al.* 1992). Instead, one could consider other mutants, partial knockdowns, or any suitable environmental manipulation strategy. Another complication is to distinguish instances of maternal or paternal imprinting. One solution would be to set up reciprocal crosses to delineate these effects (*i.e.*, perturbation of progenitor mother *vs.* father).

**Conclusion:** Our theory attempts to connect recent observations of the dynamic properties of epialleles (*i.e.*, DNA methylation variants) to a long tradition of quantitative genetics. We have shown that in the case of fully stable epialleles, genetic and epigenetic inheritance are formally indistinguishable. This illustrates that there is actually no dichotomy between these two modes of inheritance. Rather, they should be viewed as different points on a continuum that ranges from stable to unstable inheritance. We therefore hope that our work will help to bridge the gap between the fields of genetics and epigenetics.

We thank several anonymous reviewers for detailed suggestions. We also thank Ritsert C. Jansen and Vincent Colot for helpful comments on earlier versions of this manuscript. F. Johannes was supported by a Horizon Breakthrough grant (Netherlands Organisation for Scientific Research (NOW)), and M. Colomé-Tatché acknowledges support from the Centre for Quantum Engineering and Space-Time Research (QUEST).

#### LITERATURE CITED

- BIEMONT, C., 2010 Inbreeding effects in the epigenetic era. *Nat. Rev. Genet.* **11**: 234.
- BLEWITT, M., N. VICKARYOUS, A. PALDI, H. KOSEKI and E. WHITELAW, 2006 Dynamic reprogramming of DNA methylation at an epigenetically sensitive allele in mice. *PLoS Genet.* **2**: e49.
- BONDURIANSKY, R., and T. DAY, 2009 Nongenetic inheritance and its evolutionary implications. *Annu. Rev. Ecol. Evol. Syst.* **40**: 103–125.
- BORGEL, J., S. GUIBERT, Y. LI, H. CHIBA, D. SCHÜBELER *et al.*, 2010 Targets and dynamics of promoter DNA methylation during early mouse development. *Nat. Genet.* **42**: 1093–1100.
- BOSSDORF, O., C. RICHARDS and M. PIGLIUCCI, 2008 Epigenetics for ecologists. *Ecol. Lett.* **11**: 106–115.
- BOSSDORF, O., D. ARCURI, C. RICHARDS and M. PIGLIUCCI, 2010 Experimental alteration of DNA methylation affects the phenotypic plasticity of ecologically relevant traits in *Arabidopsis thaliana*. *Evol. Ecol.* **24**: 541–553.
- BROMAN, K., 2005 The genomes of recombinant inbred lines. *Genetics* **169**: 1133–1146.
- BULMER, M. G., 1985 *The Mathematical Theory of Quantitative Genetics*. Clarendon Press, Oxford.

- CASTLE, W., 1921 An improved method of estimating the number of genetic factors concerned in cases of blending inheritance. *Science* **54**: 223.
- COKUS, S., S. FENG, X. ZHANG, Z. CHEN, B. MERRIMAN *et al.*, 2008 Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452**: 215–219.
- CSABA, P., 1998 Plasticity, memory and the adaptive landscape of the genotype. *Proc. R. Soc. Lond. B* **265**: 1319–1323.
- DAXINGER, L., and E. WHITELAW, 2010 Transgenerational epigenetic inheritance: more questions than answers. *Genome Res.* **20**: 1623–1628.
- EICHLER, E., J. FLINT, G. GIBSON, A. KONG, S. LEAL *et al.*, 2010 Missing heritability and strategies for finding the underlying causes of complex disease. *Nat. Rev. Genet.* **11**: 446–450.
- FENG, S., S. COKUS, X. ZHANG, P.-Y. CHEN, M. BOSTICK *et al.*, 2010 Conservation and divergence of methylation patterning in plants and animals. *Proc. Nat. Acad. Sci. USA* **107**: 8689–8694.
- FRANKLIN, I., 1970 Average recombination frequencies. *Genetics* **66**: 709–711.
- HALDANE, J., and C. WADDINGTON, 1931 Inbreeding and linkage. *Genetics* **16**: 357–374.
- JOHANNES, F., V. COLOT and R. JANSEN, 2008 Epigenome dynamics: a quantitative genetics perspective. *Nat. Rev. Genet.* **9**: 883–890.
- JOHANNES, F., E. PORCHER, F. TEIXEIRA, V. SALIBA-COLOMBANI, M. SIMON *et al.*, 2009 Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet.* **5**: e1000530.
- KAMINSKY, Z., T. TANG, S.-C. WANG, C. PTAK, G. OH *et al.*, 2009 DNA methylation profiles in monozygotic and dizygotic twins. *Nat. Genet.* **41**: 240–245.
- KIMURA, M., 1963 A probability method for treating inbreeding systems, especially with linked genes. *Biometrics* **19**: 1–17.
- KONG, A., V. STEINTHORSDDOTTIR, G. MASSON, G. THORLEIFSSON, P. SULEM *et al.*, 2009 Parental origin of sequence variants associated with complex diseases. *Nature* **462**: 868–874.
- LANDE, R., 1981 The minimum number of genes contributing to quantitative variation between and within populations. *Genetics* **99**: 541–553.
- LI, E., T. H. BESTOR and R. JAENISCH, 1992 Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**: 915–926.
- LISTER, R., R. O'MALLEY, J. TONTI-FILIPPINI, B. GREGORY, C. BERRY *et al.*, 2008 Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**(3): 523–536.
- LIU, B. H., 1998 *Statistical Genomics*. CRC Press, Boca Raton, FL.
- LYNCH, M., and B. WALSH, 1998 *Genetics and the Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.
- MAHER, B., 2008 Personal genomes: the case of the missing heritability. *Nature* **456**: 18–21.
- MANOLIO, T., F. COLLINS, N. COX, D. GOLDSTEIN, L. HINDORFF *et al.*, 2009 Finding the missing heritability of complex diseases. *Nature* **461**: 747–753.
- MATHIEU, O., J. REINDERS, M. CAIKOVSKI, C. SMATHAJITT and J. PASZKOWSKI, 2007 Transgenerational stability of the *Arabidopsis* epigenome is coordinated by CG methylation. *Cell* **130**(5): 851–862.
- MIROUZE, M., J. REINDERS, E. BUCHER, T. NISHIMURA, K. SCHNEEBERGER *et al.*, 2009 Selective epigenetic control of retrotransposition in *Arabidopsis*. *Nature* **461**: 427–430.
- PÁL, C., and I. MIKLÓS, 1999 Epigenetic inheritance, genetic assimilation and speciation. *J. Theor. Biol.* **200**: 19–37.
- PETRONIS, A., 2010 Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature* **465**: 721–727.
- RAKYAN, V., M. BLEWITT, R. DRUKER, J. PREIS and E. WHITELAW, 2002 Metastable epialleles in mammals. *Trends Genet.* **18**(7): 348–351.
- RAKYAN, V., S. CHONG, M. CHAMP, P. CUTHBERT, H. MORGAN *et al.*, 2003 Transgenerational inheritance of epigenetic states at the murine Axin (Fu) allele occurs after maternal and paternal transmission. *Proc. Natl. Acad. Sci. USA* **100**: 2538–2543.
- REIK, W., 2007 Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447**: 425–432.
- REINDERS, J., B. WULF-BRANDE, M. MIROUZE, A. MARI-ORDONEZ, M. DAPP *et al.*, 2009 Compromised stability of DNA methylation and transposon immobilization in mosaic *Arabidopsis* epigenomes. *Genes Dev.* **23**: 939–950.
- RICHARDS, E., 2006 Inherited epigenetic variation—revisiting soft inheritance. *Nat. Rev. Genet.* **7**: 395–401.
- RICHARDS, E., 2008 Population epigenetics. *Curr. Opin. Genet. Dev.* **18**(2): 221–226.
- RICHARDS, E., 2009 Quantitative epigenetics: DNA sequence variation need not apply. *Genes Dev.* **23**: 1601–1605.
- RIESENBERG, L., M. ARCHER and R. WAYNE, 1999 Transgressive segregation, adaptation and speciation. *Heredity* **83**: 363–372.
- SEREBROVSKY, A., 1928 An analysis of the inheritance of quantitative transgressive characters. *Z. Indukt. Abstammungs. Vererbungsl.* **48**: 229–243.
- SLATKIN, M., 2009 Epigenetic inheritance and the missing heritability problem. *Genetics* **182**: 845–850.
- TAL, O., E. KISDI and E. JABLONKA, 2010 Epigenetic contribution to covariance between relatives. *Genetics* **184**: 1037–1050.
- TEIXERA, F., and V. COLOT, 2010 Repeat elements and the *Arabidopsis* DNA methylation landscape. *Heredity* **105**: 14–23.
- TEIXEIRA, F., F. HEREDIA, A. SARAZIN, F. ROUDIER, M. BOCCARA *et al.*, 2009 A role for RNAi in the selective correction of DNA methylation defects. *Science* **323**: 1600–1604.
- TSUKAHARA, S., A. KOBAYASHI, A. KAWABE, O. MATHIEU, A. MIURA *et al.*, 2009 Bursts of retrotransposition reproduced in *Arabidopsis*. *Nature* **461**: 423–426.
- VAUGHN, M., M. TANURDZIC, Z. LIPPMAN, H. JIANG, R. CARRASQUILLO *et al.*, 2007 Epigenetic natural variation in *Arabidopsis thaliana*. *PLoS Biol.* **5**: e174.
- VERHOEVEN, K., J. JANSEN, P. VAN DIJK and A. BIERE, 2010 Stress-induced DNA methylation changes and their heritability in asexual dandelion. *New Phytol.* **185**: 10.1111/j.1469-8137.2009.03121.x.
- VONGS, A., T. KAKUTANI, R. MARTIENSSEN and E. RICHARDS, 1993 *Arabidopsis thaliana* DNA methylation mutants. *Science* **260**: 1926–1928.
- WRIGHT, S., 1933 Inbreeding and homozygosity. *Proc. Natl. Acad. Sci. USA* **19**: 411–420.
- ZEMACH, A., I. MCDANIEL, P. SILVA and D. ZILBERMAN, 2010 Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328**: 916–919.
- ZENG, Z.-B., 1992 Correcting the bias of Wright's estimates of the number of genes affecting a quantitative character: a further improvement. *Genetics* **131**: 987–1001.
- ZHANG, X., S. SHIU, A. CAL and J. BOREVITZ, 2008 Global analysis of genetic and epigenetic and transcriptional polymorphisms in *Arabidopsis thaliana* using whole genome tiling arrays. *PLoS Genet.* **4**: e1000032.

APPENDIX A

This appendix shows the derivation of the epigenetic variance in the population, Equation 5. For the details on the values of the parameters presented, we refer the reader to the supporting information. The epigenetic variance can be written as

$$\sigma^2(\eta_j, t) \simeq \delta^2 N \left\{ \sigma^2(\eta_j | t) + (N - 1) \sigma(\eta_j, \eta_k | t, \bar{r}) \right\},$$

where  $\sigma^2(\eta_j | t)$  is the variance at a single locus  $j$  at time  $t$  and  $\sigma(\eta_j, \eta_k | t, \bar{r})$  is the covariance between any two loci  $j$  and  $k$  separated by an average pairwise recombination fraction  $\bar{r}$  (FRANKLIN 1970).

**Variance at a single locus:** The 12 different epigenotypes at locus  $j$  resulting from the initial cross (Figure 2A) can be classified into four different classes of three elements each. Denote these four classes by  $s_1, \dots, s_4$  with corresponding probability weights  $W_1, \dots, W_4$  (supporting material in File S1, 1.1). Since each locus can belong to any one of these classes, its variance can be written as the sum of the within-class variances weighted by their probabilities:

$$\begin{aligned} \sigma^2(\eta_j, t) &= \sum_{m=1}^4 W_m \sigma^2(\eta_j | s_m, t) \\ &= \sum_{m=1}^4 W_m \left( \sum_{i=1}^4 \mathbf{K}^{(i)}(t) \eta_{s_m}^{(i)}(t)^2 - \left( \sum_{i=1}^4 \mathbf{K}^{(i)}(t) \eta_{s_m}^{(i)}(t) \right)^2 \right), \end{aligned} \tag{7}$$

where  $\mathbf{K}(t)$  is a vector of expected single-locus epigenotype probabilities and  $\boldsymbol{\eta}(t)$  is a vector of expected epigenotypes at locus  $j$  at time  $t$  (supporting material in File S1, 1.2). The probability vector  $\mathbf{K}(t)$  is obtained following a Markov chain approach. Below we specify the cases for selfing and sibling mating.

**Selfing:**  $\mathbf{K}(t = 0)$  is a three-dimensional vector with the probabilities of each single-locus epigenotype, determined by the type of base population (F2 or BC) (supporting material in File S1, 1.3.1). The transition matrix  $\hat{T}$  for each class is a  $3 \times 3$  matrix of transition probabilities (supporting material in File S1, 1.3.2). Using a Markov chain approach we calculate the frequency of each epigenotype at time  $t$  using  $\mathbf{K}(t) = \mathbf{K}(0) \hat{T}^t$ . Using Equation 7 the variance at a single locus is given by

$$\sigma^2(\eta_j, t) = \frac{1}{4} \{ q_1 [4\gamma(t)(\gamma(t) + 1)(s - 1) - 3s - 1] \}. \tag{8}$$

The value of the parameter  $q_1$  differs for the type of base population considered (F2 or BC) and for the type of epigenotypic effect considered (additivity or dominance); see Table A1.

**Sibling mating:** Consider, instead of the probabilities of each single-locus, the probabilities of their mating types (BULMER 1985). There are 16 possible mating types in each class, which can be reduced to 6 consider-

ing the following basic symmetries (supporting material in File S1, 1.4.1):  $\Omega \cdot \tilde{\omega} = \tilde{\omega} \cdot \Omega$ , and  $\Omega \cdot \tilde{\omega} \times \Omega \cdot \Omega = \Omega \cdot \Omega \times \Omega \cdot \tilde{\omega}$ . The transition matrix  $\hat{T}$  is the collection of the probabilities of going from one mating type to another in one generation of sibling mating; it has dimension  $6 \times 6$  (supporting material in File S1, 1.4.3) (for a detailed description on how to construct such a matrix, see BULMER 1985, Chap. 3). The initial probabilities of each mating class,  $\mathbf{Q}(0)$ , is a 6-dimensional vector given by  $Q_{(l)}(0) = K_{(i)}(0)K_{(j)}(0)$ , where  $i$  and  $j$  are the single-locus epigenotypes involved in mating type  $l$ . At any generation  $t$ ,  $\mathbf{Q}(t) = \mathbf{Q}(0) \hat{T}^t$ . The probabilities for the 4 different single-locus epigenotypes are given by  $K_{(i)}(t) = \sum_{l=1}^6 p_{il} Q_{(l)}(t)$ , where  $p_{il}$  is the proportion of single-locus  $i$  involved in mating type  $l$  (supporting material in File S1, 1.4.2). Using Equation 7 the variance at a single locus is given by

$$\sigma^2(\eta_j, t) = \frac{1}{4} \{ q_1 [4\gamma(t)(\gamma(t) + 1)(s - 1) - 3s - 1] \}. \tag{9}$$

The value of the parameter  $q_1$  depends on the case considered (Table A2).

**Covariance between loci:** To calculate the covariance term between loci  $j$  and  $k$  separated by an average recombination factor  $\bar{r}$  we use a similar classification for the possible 160 different two-locus epigenotypes resulting from the initial cross (Figure 2A). They can be assigned to 16 different classes of 10 pairs each,  $d_1, \dots, d_{16}$ , with probability  $V_i(\eta_j, \eta_k)$  (supporting material in File S1, 2.1). Since each two-locus epigenotype can belong to any one of these 16 classes we can write the covariance as:

$$\begin{aligned} \sigma(\eta_j, \eta_k | t, \bar{r}) &\simeq \sum_{m=1}^{16} V_m(\eta_j, \eta_k) \sigma(\eta_j, \eta_k | d_m, t, \bar{r}) \\ &= \sum_{m=1}^{16} V_m(\eta_j, \eta_k) \left[ \sum_{i=1}^{10} R^{(i)}(t) \eta_{d_{m,1}}^{(i)}(t) \eta_{d_{m,2}}^{(i)}(t) \right. \\ &\quad \left. - \left( \sum_{i=1}^{10} R^{(i)}(t) \eta_{d_{m,1}}^{(i)}(t) \right) \left( \sum_{i=1}^{10} R^{(i)}(t) \eta_{d_{m,2}}^{(i)}(t) \right) \right], \end{aligned} \tag{10}$$

where  $\mathbf{R}(t)$  is a 10-dimensional vector of two-locus epigenotype probabilities (*i.e.*, parental haplotypes). Here,  $\eta_{d_{m,1}}$  and  $\eta_{d_{m,2}}$  are the two different expected epigenotypes in a given class  $d_m$  (supporting material in File S1, 2.2). The probability vector  $\mathbf{R}(t)$  is obtained following a Markov chain approach; below we specify the cases for selfing and sibling mating. In the simplified case of purely Mendelian inheritance ( $s = 1$ ), no transgression ( $\tau = 0$ ), and  $t = \infty$ , the derivation of  $\mathbf{R}(t_\infty)$  has received considerable attention as a problem in its own right (HALDANE and WADDINGTON 1931; WRIGHT 1933; KIMURA 1963; BROMAN 2005).

**Selfing:** For each of the 16 classes  $\mathbf{R}(t = 0)$  is calculated depending on the type of base population considered (F2 or BC) (supporting material in File S1, 2.3.1). The transition probability matrix  $\hat{T}$  is a  $10 \times 10$  matrix of

TABLE A1

Base	Effect	$q_1$
F2	Add.	$\frac{1}{2^{t+1}} - 1$
F2	Dom.	$\frac{1 - 2^{2+2t}}{2^{2t+2}}$
BC	Add.	$\frac{1}{2^{t+1}} - \frac{3}{4}$
BC	Dom.	$\frac{1 + 2^{t+1}\xi - 3 \times 2^{2t}}{2^{2t+2}},$ $\xi = 1 \rightarrow$ wt dominant, $\xi = -1 \rightarrow$ mutant dominant

transition probabilities of each two-locus epigenotype crossed with itself (supporting material in File S1, 2.3.2), and  $\mathbf{R}(t) = \mathbf{R}(0)\hat{T}^t$ . Using Equation 10 we obtain

$$\sigma(\eta_j, \eta_k | t, \bar{r}) \simeq \frac{1}{4} \left\{ q_2 [1 + s + 2\gamma(1 - s)]^2 (1 - 2\tau)^2 \right\}. \quad (11)$$

The analytical value of the parameter  $q_2$  is shown in Table A3.

**Sibling mating:** Consider the 55 different mating types between the 10 two-locus epigenotypes in each class (supporting material in File S1, 2.4.1) (BULMER 1985). Taking into account the symmetries mentioned above we can reduce them to 22 different mating types for the F2-generated population and to 34 for the BC one. In the same way as for the single-locus case,  $\mathbf{Q}(0)$  are the initial probabilities of each mating class,  $\mathbf{Q}(t) = \mathbf{Q}(0)\hat{T}^t$ , and the probabilities for the 10 two-locus epigenotypes,  $\mathbf{R}(t)$ , can be extracted from the mating type probabilities  $\mathbf{Q}(t)$  (supporting material in File S1, 2.4.2).

Unfortunately, the transition matrix  $\hat{T}$  cannot be diagonalized symbolically because its dimension is too large (see Figure S2). For this reason we cannot obtain an analytical expression (as a function of the parameter  $\bar{r}$ ) for the probabilities  $\mathbf{R}(t)$ . However, we can fix  $\bar{r}$  to a numerical value before calculating the power  $\hat{T}^t$ , and thus obtain the exact probability values at any time  $t$ . Moreover, it is possible to write symbolically the probability vector as  $\mathbf{R}(t) = \{R^{(1)}(t), R^{(2)}(t), \dots, R^{(10)}(t)\}$  and, using Equation 10 and taking into account  $\sum_{i=1}^{10} R_{(i)} = 1$ , we can write the covariance for sibling mating in the form

$$\sigma(\eta_j, \eta_k | t, \bar{r}) \simeq \frac{1}{4} \left\{ q_2 [1 + s + 2\gamma(1 - s)]^2 (1 - 2\tau)^2 \right\}, \quad (12)$$

where the constant  $q_2$  is calculated exactly for any value of  $\bar{r}$  for a F2- or a BC-based population.

### Epigenetic variance:

Finally, combining Equations 8 and 9 with Equations 11 and 12, and multiplying by the number of loci and their mean phenotypic effect,  $N\delta^2$ , yields Equation 5 in the main text.

### APPENDIX B

In this appendix we show how we estimate the number of QTL in the *ddm1*-derived epiRILs. Consider the equation for  $N$  in the main text (Equation 6) at  $t = \infty$  (*i.e.*, fully inbred), under the assumption of perfect resetting ( $\gamma(t = \infty) = -\frac{1}{2}$ ):

$$N = \frac{3D^2 s (s(1 - 2\tau)^2 + (2\bar{r} + 1)/(2\bar{r} - 1))}{(3D^2 s^2 + 16\sigma^2(\eta, t)(2\bar{r} + 1)/(2\bar{r} - 1))(1 - 2\tau)^2} = f(\Psi, t),$$

where  $\Psi$  is a vector of all the parameters specified on the right-hand side of the equation. Substituting the

TABLE A2

Base	Effect	$q_1$
F2	Add.	$-1 - \frac{1}{4} \left( \frac{1 - \sqrt{5}}{4} \right)^t \left( \frac{3}{\sqrt{5}} - 1 \right) + \frac{1}{4} \left( \frac{1 + \sqrt{5}}{4} \right)^t \left( \frac{3}{\sqrt{5}} + 1 \right)$
F2	Dom.	$\frac{1}{5} \left( \frac{(-1)^{t+1}}{2^{1+2t}} - \frac{1}{2^{3+4t}} \left[ (3\sqrt{5} - 7)(1 - \sqrt{5})^{2t} - (3\sqrt{5} + 7)(1 + \sqrt{5})^{2t} \right] - 5 \right)$
BC	Add.	$-\frac{3}{4} - \frac{1}{4} \left( \frac{1 - \sqrt{5}}{4} \right)^t \left( \frac{2}{\sqrt{5}} - 1 \right) + \frac{1}{4} \left( \frac{1 + \sqrt{5}}{4} \right)^t \left( \frac{2}{\sqrt{5}} + 1 \right)$
BC	Dom.	$\frac{1}{5} \left( \frac{(-1)^t}{2^{3+2t}} - \frac{1}{2^{2+2t}} \left[ \xi(5 - 2\sqrt{5})(1 - \sqrt{5})^t + \xi(5 + 2\sqrt{5})(1 + \sqrt{5})^t \right] + \frac{1}{2^{2+2t}} \left[ (9 - 4\sqrt{5})(1 - \sqrt{5})^{2t} + (9 + 4\sqrt{5})(1 + \sqrt{5})^{2t} \right] - \frac{15}{4} \right),$ $\xi = 1 \rightarrow$ wt dominant, $\xi = -1 \rightarrow$ mutant dominant

TABLE A3

Base	Effect	$q_2$
F2	Add.	$\frac{2\bar{r}-1}{2\bar{r}+1} \left( \frac{(1-2\bar{r})^{t+1}}{2^{t+1}} - 1 \right)$
F2	Dom.	$\frac{1}{2^{t+1}} \left( (1+2\bar{r}(\bar{r}-1))^t \frac{(1-2\bar{r}^2+4\bar{r}^3)}{1+2\bar{r}} - \frac{(1-2\bar{r})^{t+2}}{1+2\bar{r}} \right) + \frac{1-2\bar{r}}{1+2\bar{r}} - \frac{1}{2^{t+2}}$
BC	Add.	$\frac{2\bar{r}-1}{2\bar{r}+1} \left( \frac{(1-2\bar{r})^t(1-\bar{r})}{2^{t+1}} - \frac{3}{4} \right)$
BC	Dom.	$-\frac{1}{2^{t+1}} \left( \frac{\xi(1-4\bar{r}^2)}{2\bar{r}+1} + (\bar{r}-1) \left[ (1+2\bar{r}(\bar{r}-1))^t - \frac{(1-2\bar{r})^{t+1}}{1+2\bar{r}} \right] \right) - \frac{3}{4} \frac{2\bar{r}-1}{2\bar{r}+1} - \frac{1}{2^{t+2}}$

$\xi = 1 \rightarrow$  wt dominant,  $\xi = -1 \rightarrow$  mutant dominant

values for  $D$ ,  $\sigma^2(\eta, t)$ ,  $s$ , and  $\bar{r}$  provided in the main text yields one equation and two unknowns ( $N$  and  $\tau$ ). With heritability data from only one generation, it is not possible to find unique solutions, and at least one additional generation of phenotypic measurements is required.

**Obtaining  $\bar{N}$ :** In the absence of such additional data one strategy is to calculate an average  $N$  by integrating over the theoretical range of  $\tau$ ,

$$\bar{N} = \frac{\int_0^u f(\Psi, t) d\tau}{\int_0^u d\tau},$$

where  $u = \tau|(N = N_{\max})$  is the upper integration limit. To find a value for  $u$  we solve Equation 5 for  $\tau$  and evaluate it at the expected maximum number of QTL,  $N_{\max}$ , which can be detected given the particular mating scheme.

**Obtaining the expected value for  $N_{\max}$ :** In a population of RILs, let  $R_j$  denote the probability of a recombinant type over the entire length of the  $j$ th chromosome,  $R_j = \sum_{i \in \mathcal{R}} R^{(i)}(t = \infty)$ , where  $R^{(i)}(t)$  are the components of  $\mathbf{R}(t)$  (Appendix A) at fixation and  $\mathcal{R}$  is the ensemble of recombinant two-locus epigenotypes. Using our Markov chain approach for selfing, the value of  $R_j$  is given by

$$R_j|F2 = \frac{2r_j}{(1+2r_j)}, \quad R_j|BC = \frac{3r_j}{(2+4r_j)}, \quad (13)$$

where  $r_j$  is the recombination fraction at meiosis between the beginning and the end of chromosome  $j$ . Note that the result for  $R_j|F2$  is consistent with HALDANE and WADDINGTON (1931). Given a known genetic map,  $r_j$  can be calculated using any map function, as long as its inverse is available (LIU 1998). The probability of a recombinant type implies at least one recombination breakpoint in the interval, thus generating two potential QTL segments flanking the breakpoint. Assuming  $s = 1$  (all epialleles are stable), the expected maximum num-

ber of QTL occurs in a situation where each generated segment is occupied by a QTL. This expectation can be approximated by

$$E(N_{\max}) \simeq 2 \sum_{j=1}^C \frac{R_j}{1-R_j},$$

where  $C$  is the total number of chromosomes, and the ratio on the right-hand side is the odds ratio of a recombination *vs.* no recombination breakpoint on chromosome  $j$ . As a rule of thumb, it is safe to say that  $E(N_{\max}|F2) \simeq 2C$  and  $E(N_{\max}|BC) \simeq \frac{6}{5}C$ . These latter expressions assume linkage equilibrium between the beginning and end of chromosome  $j$ , that is  $r_j = 0.5$ .

**Bootstrap standard errors:** We obtain standard errors for  $\bar{N}$  using a nonparametric bootstrap approach. To achieve this we take the following steps:

1. Recalculate  $D$  on the basis of a random sample of size  $n$  from each of the two parental phenotypic vectors.
2. Draw a random, stratified bootstrap sample from the epiRILs phenotypic vector, and approximate the epigenetic variance,  $\sigma^2(\eta, t)$ , using a random intercepts model:  $y_{i,j} = \beta_0 + b_{i,z_{i,j}} + \varepsilon_{i,j}$  where  $\beta_0$  is a common fixed intercept,  $b_i$  is the random intercept of the  $i$ th line,  $z_{i,j}$  is an index variable, and  $\varepsilon_{i,j}$  is the error. We assume that  $b_i \sim N(0, \sigma^2(\eta, t = \infty))$ ,  $\varepsilon_{i,j} \sim N(0, \sigma^2_i)$ , and  $\text{Cov}(\varepsilon_{i,j}, \varepsilon_{i,j'}) = 0$ .
3. Use the estimates for 1 and 2, and determine the upper integration limit ( $u$ ) for  $\tau$ .
4. Determine  $\bar{N}$  by calculating  $\bar{N} = \int_0^u f(\Psi, t) d\tau / \int_0^u d\tau$ .
5. Repeat steps 1–4 a large number of times. The standard deviation of the resulting bootstrap distribution is an approximation for the standard error of  $\bar{N}$ .

Note that these sampling errors will be slightly underestimated, because the values for  $s$  and  $\gamma(t)$  are assumed known from molecular analysis, and the variation in  $N_{\max}$  is also neglected for simplicity.