Combinatorial Methods for Epistasis and Dominance

SERGE SVERDLOV and ELIZABETH THOMPSON

ABSTRACT

We develop computational tools for the analysis of nonlinear genotype–phenotype relationships with epistasis among multiple loci or dominance interactions among multiple alleles within the same locus. Theory distinguishes between separable traits, with removable epistasis, and traits with essential epistasis. Separable traits can be transformed to a natural scale where additive methods apply. The methods we present solve for the natural scale, exactly when possible and approximately when not. Through graph methods, our methods allow for enumeration, counting, or sampling of distinct trait architectures satisfying constraints from the separability theory. A tool is provided for diagnosing which separability constraints are violated by a given nonseparable architecture. For genetic traits controlled by limited numbers of loci and alleles, our algorithm enumerates all possible trait structures and finds exact or error-minimizing linearizing transformations by formulating a constrained optimization program. We find that the fraction of possible distinct genetic traits satisfying simple criteria that can be fully or approximately linearized is high for small systems and falls as the number of alleles or loci increases.

Keywords: combinatorics, epistasis, genotype-phenotype map, integer programming, linear programming, population genetics, quantitative genetics, separability.

1. INTRODUCTION: NONLINEARITY AND THE NATURAL SCALE

1.1. Linear and nonlinear models

MANY MODELS WITHIN QUANTITATIVE GENETICS are based on the assumption of linearity or additivity, where individual alleles make additive contributions to a quantitative trait or a latent risk. Deviations from such additivity take the form of epistasis, or nonlinear interaction of contributions from multiple genes or loci, and dominance, nonlinearity within a diploid gene.

The theory of nonlinearity in quantitative genetics as developed in Sverdlov and Thompson (In Prep) distinguishes between separable traits, where a *natural scale* can be found on which the trait can be treated as additive, and traits possessing essential epistasis, for which no such scale exists. Consider a quantitative trait T controlled by a single diploid locus:

$$E[T] = \mu + \alpha_i + \alpha_j + \delta_{ij}.$$

Here, the expectation is taken over possible environments, μ is an optional centering parameter and α_i is the additive effect associated with allele i . If it is possible to assign values α_i so as to make $\delta_{ij} = 0$ for all i and j , the trait is additive. If the trait is not additive, but there is a transformation $f(\cdot)$ such that

$$E[f(T)] = \mu^{(f)} + \alpha_i^{(f)} + \alpha_j^{(f)} + \delta_{ij},$$

with all $\delta_{ij}^{(f)} = 0$, then the trait is *separable*, and the scale to which $f(\cdot)$ transforms T is the *natural scale*. Additivity can be expressed in matrix notation; if we form the matrix of expected trait values $\mathbf{T}_{ij} = E[f(T_{ij})] = \alpha_i + \alpha_j$ for some vector α , then

$$\mathbf{T} = \alpha' \mathbf{1} + \mathbf{1}' \alpha.$$

Diploid symmetry ensures that $\mathbf{T}_{ij} = \mathbf{T}_{ji}$.

For the epistasis case, the phenotype in an additive model is the sum of contributions from L loci, with symmetric contributions for the paired alleles within each locus,

$$E[T] = \sum_{l=1}^L (\alpha_{(l)i} + \alpha_{(l)j}).$$

As in the single locus diploid case, for a general trait T such a decomposition does not exist, but we are interested in a transformation to the natural scale $f(\cdot)$ for which a choice of $\alpha^{(f)}$ exists such that

$$E[f(T)] = \sum_{l=1}^L (\alpha_{(l)i}^{(f)} + \alpha_{(l)j}^{(f)}).$$

1.2. Ordinal scale

Since we are interested in transforming traits to a natural scale, the original trait's given scale can be abstracted away. Consider the ordinal scale transformation $f(T_{ij}) \rightarrow \text{rank}(T_{ij}, \{T_{11}, T_{12}, \dots, T_{nn}\})$, which assigns 1 to the lowest trait value, 2 to the next higher trait value, and so on. Such a representation is as informative for the purpose of finding an original scale as the original representation. A trait architecture can be specified up to scale transformation by a complete ordering of the genotypes; an incomplete (but consistent) ordering specifies a set of trait architectures.

Our method is based on the ordering of genotypes according to their genotypic values, that is, according to the expectations of phenotypes over possible environments. A scale transformation does not always preserve order under expectation. Complications due to this phenomenon are considered in Sverdlov and Thompson (In Prep); for the purposes of computational methods developed here, we assume no environmental variability, $E T = T$ and $E f(T) = f(T)$, which ensures an order invariant to monotonic transformations.

1.3. Computational methods

In this article, we introduce a combinatorial approach to dominance and epistasis. Ideally, the question of whether epistasis or dominance exists as a biological phenomenon should be considered by a method invariant to monotonic transformation. This leads us to consider ordinal, rather than quantitative, traits. By specifying traits as an ordering relation or ranking over genotypes, we can ask questions about traits without reference to the scale on which they are measured. Can we choose a *natural scale* on which the trait is additive, that is, on which dominance and epistasis do not exist? Under what assumptions can such a scale be chosen exactly or approximately? This combinatorial approach contrasts with methods for selecting a phenotype scale by transformation to a normal phenotype distribution (e.g., Fusi et al., 2014).

We apply the combinatorial approach with similar methods to both dominance and epistasis. We demonstrate the role of *directional consistency of substitution* assumption and illustrate the combinatorially large number of discrete epistatic architectures, which may be classified according to biologically relevant properties.

We introduce a variety of computational tools for dealing with both separable and approximately separable traits. We develop natural scale recovery as constrained optimization problem for both the diploidy (dominance) and epistasis cases. The optimization finds an exact transformation to the natural scale if it exists and minimizes the number of ordering constraints that must be relaxed to find a solution if no exact solution exists. We develop tools to enumerate, count, or sample ordinal trait architectures satisfying certain constraints necessary but not sufficient for additivity. For the epistatic case, we develop a tool to find features within the trait architecture that explain the impossibility of finding a natural scale.

2. DIPLOID TRAITS AND DOMINANCE

2.1. Diploid constraints and directional consistency of substitution

In the single locus, diploidy with dominance, context, we can identify several implications of additivity, which have a biological interpretation but are invariant to a monotonic transformation of the trait's scale, in particular *Directional Consistency of Substitution* and *No Overdominance*.

In Section 1.1, we described the conditions under which the trait matrix $\mathbf{T}_{ij}$ represents an additive trait. These conditions are implied by separability and therefore necessary for it, but not sufficient.

Consider a single diploid locus with alleles A, B, C, D . We will refer to four alleles labeled by letters for convenience of exposition; the generalization to an arbitrary number of alleles is straightforward. Each pair of alleles is associated with a trait value, and we are given an ordering relation $\succeq$ on those value. It will be convenient to assume that each genotype has a unique trait value; that is, that the genotypes can be ranked with no ties, forming a complete order relation $\succ$. We will write, for example, $AB \succ AC$ to denote that AB has trait value higher than genotype AC . Assume a complete ordering; that is, all trait values are distinct (e.g., either $AD \succ BC$ or $AD \prec BC$ but not $AD \approx BC$), except matching heterozygotes ($AB \approx BA$).

Without loss of generality, we order the allele labels alphabetically according to their homozygotes,

$$AA \prec BB \prec CC \prec DD.$$

The principle of nonoverdominance is that for any P, Q such that $PP \prec QQ$,

$$PP \prec PQ \prec QQ.$$

By using strict inequalities, we also rule out complete dominance ($PQ \approx QQ$); this is consistent with the assumption of a distinct value for each genotype.

We introduce the principle of directional consistency of substitution (DCS): substituting an allele R for Q has either a positive or a negative effect, regardless of the identity of the unchanged allele. That is,

$$PQ \succ PR \rightarrow SQ \succ SR.$$

DCS implies nonoverdominance because either $PP \succ PQ$ implying $QP \succ QQ$ or $PQ \succ PP$ implying $QQ \succ QP$. We can say DCS generalizes the idea of nonoverdominance to multiple alleles.

DCS implies an ordering on the individual alleles A, B, C that is the same as the ordering of homozygote genotypic values. If $PP \prec QQ$, then substitution of Q for P has a positive effect.

Under DCS, the matrix of genotypic values expressed on any scale

$$\mathbf{T} = \begin{bmatrix} T_{AA} & T_{AB} & T_{AC} & T_{AD} \\ T_{AB} & T_{BB} & T_{BC} & T_{BD} \\ T_{AC} & T_{BC} & T_{CC} & T_{CD} \\ T_{AD} & T_{BD} & T_{CD} & T_{DD} \end{bmatrix},$$

is monotonically increasing both along its rows and along its columns. Directional consistency and non-overdominance can be expressed as partial orderings on the set of genotypes (given a fixed order of homozygotes), as illustrated in Figure 1.

2.2. Linearization and relaxed linearization

Both DCS and No Overdominance are implied by additivity and therefore necessary for the trait to be additive on some scale. From Figure 1, we immediately see that directional consistency implies No Overdominance and is therefore the stricter condition. The broader question is under what circumstances are these conditions sufficient for the existence of a natural scale. We can pose two forms of the natural scale problem. In the strict formulation, given a diploid trait with genotypic values T_{ij} on some arbitrary scale (e.g., on a rank scale from 1 to $\binom{n}{2} + \binom{n}{1} = \frac{n^2 + n}{2}$, the number of distinct genotypes for n alleles), is there a *strictly increasing* function f such that there exists $\alpha_1 \dots \alpha_n$ for which

FIG. 1. Ordering relation constraints for single diploid locus. Left: no overdominance. Right: DCS. DCS, directional consistency of substitution.

$$f(T_{ij}) = \alpha_i + \alpha_j.$$

Counterexamples in Sverdlov and Thompson (In Prep) illustrate that a transformation to additivity is not always possible, even given directional consistency. A relaxed formulation of the problem is as follows: does there exist such a function f that is *nondecreasing*, but not trivially a constant? The constant valued function $f(\cdot) = c$ is always a nondecreasing solution; a nontrivial solution would vary over the domain of $f(\cdot)$.

2.3. Traversal trees

The ordering constraint on the genotypes of a single locus system imposed by nonoverdominance or DCS, as in Figure 1, forms a directed acyclic graph (DAG), with genotype nodes and edges corresponding to each given constraint of the form $AA \prec AB$. A fully specified trait corresponds to a complete ordering, and its DAG consists of a single connected path. Fixing the order of homozygotes with no loss of generality as $AA \prec BB \prec CC \prec DD$, we can describe certain ordering constraints by an *ordering constraint DAG* as in Figure 1. The DAG corresponding to an ordering constraint is not generally unique. In particular, we can perform the transitive reduction (Aho et al., 1972) and transitive closure (Purdom, 1970) operations to find the minimal and maximal DAG, respectively. Equivalence of two ordering constraint DAGs is then given by the equivalence of their transitive closures. A fully specified trait corresponds to a complete ordering, and its DAG consists of a single connected path. Fixing the order of homozygotes with no loss of generality as $AA \prec BB \prec CC \prec DD$, we can describe certain ordering constraints by an *ordering constraint DAG* as in Figure 1. The DAG corresponding to an ordering constraint is not generally unique. In particular, we can perform the transitive reduction Aho et al. (1972) and transitive closure Purdom (1970) operations to find the minimal and maximal DAG, respectively. Equivalence of two ordering constraint DAGs is then given by the equivalence of their transitive closures.

Any complete traversal of this DAG is an ordering of the nodes (genotypes) consistent with the ordering constraint, and therefore, a trait architecture. For example,

$$AA \prec AD \prec AC \prec AB \prec BB \prec BC \prec BD \prec CC \prec CD \prec DD,$$

is an ordering that satisfies nonoverdominance, but not DCS (e.g., $AC \prec AB$ but $BB \prec BC$).

We developed an algorithm to transform the ordering constraint DAG to the *Path DAG*. This is a DAG with a single root and a single leaf. Each path from the root to the leaf corresponds to a traversal of the original DAG, with edges of the Path DAG corresponding to nodes of the original DAG (genotypes), and nodes of the Path DAG labeled with the number of distinct paths from that node to the leaf. This allows the enumeration, and therefore, counting and equally weighted sampling, of the paths through the original DAG. Figures 2 and 6 demonstrate the original and Path DAG for DCS for single locus diploid systems

FIG. 2. Dominance patterns for $n=2, 3, 4$. Graph on the left is the allele ordering imposed on genotypes by DCS. Graph on the right is the Path DAG (Section 2.3). DAG, directed acyclic graph.

with between 2 and 7 alleles. The growth in number of directionally consistent orderings can be observed from the labeled number of paths from the root of the Path DAG.

The algorithm is analogous to the problem of generating, as in Pruesse and Ruskey (1994), or counting, as in Brightwell and Winkler (1991), the linear extensions in a DAG. Given an input DAG with allele set A and edges $a \rightarrow b$ representing ordering constraints of the form $a \prec b$, the Path DAG we will generate will have nodes corresponding to subsets of A :

FIG. 3. Patterns of dominance, up to label permutation, for 2, 3, and 4 alleles. Solvable systems are in green; partially solvable in red.

1. Initialize the Path DAG to a single node, the empty set $\emptyset$.
2. Choose a node S in the Path DAG, and an allele a not in S , such that for every $s \in S$, the edge $a \rightarrow s$ is in the input DAG.
 - (a) Add the node $S \cup \{a\}$ to the Path DAG if it does not already exist.
 - (b) Add an edge from that node to S in the Path DAG; label that edge a .
3. Repeat while such choices are possible.
4. Trim the DAG to include only paths that lead from A to $\emptyset$.
5. Assign to each node the number of paths the source node to $\emptyset$ by recursively summing the path counts of its children with memoization. Assign path count 1 to the node $\emptyset$.

Any path from A to $\emptyset$ is then a distinct trait architecture, with the order found by collecting alleles from edge labels during the traversal. The path count assigned to node A is then the overall count of distinct trait architectures.

2.4. Natural scale recovery

Given a genotype ordering, for example

$$AA \prec AB \prec BB \prec AC \prec BC \prec CC$$

we can find numerical values for a natural scale, exactly or approximately, by a linear program. We will introduce three sets of variables:

1. Genotype variables: AA, AB, BB, AC, BC, CC
Each represents the genotypic value (e.g., median phenotype) corresponding to each genotype on a particular scale.
2. Allele variables: A, B, C
Each represents the additive contribution of an allele under the assumption of additivity; for example $A + B = AB$.
3. Gap variables: $g_{AA}^{AB}, g_{AB}^{BB}, g_{BB}^{AC}, g_{AC}^{BC}, g_{BC}^{CC}$

One gap variable is introduced for each nonredundant inequality in the specified ordering. A gap variable is binary, taking on the values 0 or 1, depending on whether the inequality is satisfied strictly or nonstrictly. These variables are used both in the objective and in constraints to design a linear program that seeks an additive solution where all inequalities are satisfied in the strict sense, but will fall back to an imperfect solution.

The linear program then has the form

maximize $g_{AA}^{AB} + g_{AB}^{BB} + g_{BB}^{AC} + g_{AC}^{BC} + g_{BC}^{CC}$ subject to

1. Domain: $A, B, C \geq 0$ (can be simplified to $A=0$)
2. Linearity:

$$AA = A + A, AB = A + B, BB = B + B, AC = A + C, BC = B + C, CC = C + C$$

3. Gap variables: $0 \leq g_{AA}^{AB}, g_{AB}^{BB}, g_{BB}^{AC}, g_{AC}^{BC}, g_{BC}^{CC} \leq 1$
4. Order:

$$AB - AA \geq g_{AA}^{AB}, BB - AA \geq g_{AB}^{BB}, AC - BB \geq g_{BB}^{AC}, BC - AC \geq g_{AC}^{BC}, CC - BC \geq g_{BC}^{CC}$$

This “accordion” linear program maximizes the number of gaps of the form g_{AA}^{AB} , assigning the value of 1 to each strict inequality satisfied and 0 to each inequality relaxed into an equality. If it can, the program will make all inequalities strict; at worst, it will turn all inequalities into equalities, producing the trivial constant solution. The use of linear programming and optimization in translating order constraints to a cardinal scale can be traced to Kruskal (1964) and the development of the nonmetric version of the multidimensional scaling framework.

2.5. Solvability of dominance systems

We enumerated all possible single locus systems for 2, 3, and 4 alleles, up to permutation, and displayed the solutions as 3. For up to 4 alleles, all but two possible order architectures are solvable, that is, they have a natural scale that respects the strict inequalities of the ordering constraints. The two partially solvable orderings have relaxed or partial solutions; that is, they can be solved by turning some, but not all, of the strict inequalities into nonstrict inequalities. We tabulated results for higher numbers of alleles as Table 1. As the number of alleles rises, the fraction of architectures that either has relaxed solutions or no solutions increases.

3. MULTILOCUS TRAITS AND EPISTASIS

3.1. Role of epistasis

The relative importance of epistasis, nonlinearity over multiple loci, in explaining the variability of quantitative traits remains controversial. A survey of the varied biological and statistical concepts of epistasis is given by Phillips (2008). In particular, we must distinguish between the noncontroversial assertion that epistasis in the biological sense, at the level of genes interacting to produce the phenotype of an individual, is common and the controversy around the relative importance of epistasis as a statistical phenomenon useful in describing variation at the population level. There is an ongoing debate between the

TABLE 1. SOLVABILITY OF DOMINANCE SYSTEMS FOR $N=1 \dots 6$

N	Total	Solvable	Partially	Nonsolvable
1	1	1	0	0
2	1	1	0	0
3	2	2	0	0
4	12	10	2	0
5	286	114	132	40
6	33,592	2608	13,488	17,496

advocates (Zuk et al., 2012) and critics (Hill et al., 2008) of an important role for epistatic genetic variance in general complex traits. The canonical example of a human anthropometric trait thought to have high epistatic genetic variance is the Total Dermal Ridge Count (Heath et al., 1984), a property of fingerprints. Fingerprints are typically nearly identical in monozygotic twins, but similarities between the fingerprints of other relatives do not follow a simple pattern. Thus, total genetic variance is high but additive variance is low, pointing to a big role for epistasis.

3.2. Directional consistency in epistasis

The concepts of the dominance framework developed above carry over directly to the case of epistasis over multiple loci, but the number of combinatorial cases and both the computational and notational difficulties increase considerably.

FIG. 4. Order constraints on epistatic systems with weak directional consistency for 1, 2, or 3 loci. Green indicates ordering of loci imposed by label permutation.

FIG. 5. Order constraints on epistatic systems with weak directional consistency for 4 loci. Green indicates ordering of loci imposed by label permutation. Other colors arbitrary, for contrast only.

FIG. 6. Dominance patterns for $n=5$.

directional consistency, but neither converse holds. Weak directional consistency as an ordering constraint can be represented as a DAG over genotypes for a fixed ordering of single “uppercase” locus genotypes, but strong directional consistency cannot. Directional consistency, in a relatively weak or strong form, is still a weaker and more biologically plausible hypothesis about a given trait than additivity. We plot the order constraints due to weak directional consistency as Figures 4 and 5.

3.3. Algorithmic solutions

We developed an algorithm to solve epistatic order problems for natural scale, exactly or approximately, by linear programming, as well as to detect violations of directional consistency. For selected orderings, the output is illustrated as Fig. 9. As an example showing how to read this table, take the second block. The first row is an ordering of haploid genotypes in increasing order from *abcd* to *ABCD*. A strict solution, with no relaxation of strict inequalities, is impossible, but the maximum number of inequalities is satisfied in the best nontrivial solution, assigning $A=1, B=2, C=3, D=6$. The last row specifies the violation of directional consistency in the order given: conditioning on either *d* or *D*, the ordering of *ABc* and *abC* flips, a violation of third order consistency.

The linear program is directly analogous to that used with the diploid (dominance) systems, with locus effect variables in place of allele effect variables. To find directional consistency violations, we enumerate the possibilities for a given ordering of genotypes:

```

abcd Abcd aBcd ABcd abCd AbCd aBCd ABCd abcd Abcd aBcd ABcd abCD AbCD aBCD ABCD
strict: [1, 2, 4, 8]

abcd Abcd aBcd ABcd abCd AbCd aBCd ABCd abcd Abcd aBcd abCD ABcd AbCD aBCD ABCD
strict: Infeasible
relaxed: [1, 2, 3, 6]
abcd Abcd aBcd ABcd abCd AbCd aBCd ABCd abcd Abcd aBcd abCD ABcd AbCD aBCD ABCD
{'~~~d': ['ABc', 'abC'], '~~~D': ['abC', 'ABc']}

abcd Abcd aBcd ABcd abCd AbCd aBCd abcd Abcd ABCd aBcd ABcd abcd Abcd aBCD ABCD
strict: Infeasible
relaxed: [0, 1, 2, 3]
abcd Abcd aBcd ABcd abCd AbCd aBCd abcd Abcd ABCd aBcd ABcd abcd Abcd aBCD ABCD
{'A~~~': ['bcd', 'BCd'], 'a~~~': ['BCd', 'bcd']}

abcd Abcd aBcd ABcd abCd AbCd aBCd abcd Abcd ABCd aBcd abcd ABcd Abcd aBCD ABCD
strict: Infeasible
relaxed: [1, 2, 3, 5]
abcd Abcd aBcd ABcd abCd AbCd aBCd abcd Abcd ABCd aBcd abcd ABcd Abcd aBCD ABCD
{'A~~~': ['bcd', 'BCd'], 'a~~~': ['BCd', 'bcd']}
{'~~~d': ['ABc', 'abC'], '~~~D': ['abC', 'ABc']}

abcd Abcd aBcd ABcd abCd AbCd aBCd abcd Abcd aBcd abCD ABCd ABcd Abcd aBCD ABCD
strict: Infeasible
relaxed: Infeasible
abcd Abcd aBcd ABcd abCd AbCd aBCd abcd Abcd aBcd abCD ABCd ABcd Abcd aBCD ABCD
{'A~~~': ['bcd', 'BCd'], 'a~~~': ['BCd', 'bcd']}
{'~B~~': ['acD', 'ACd'], '~b~~': ['ACd', 'acD']}
{'~~c~': ['ABd', 'abD'], '~~C~': ['abD', 'ABd']}
{'~~~d': ['ABc', 'abC'], '~~~D': ['abC', 'ABc']}

abcd Abcd aBcd ABcd abCd AbCd aBCd abcd Abcd aBcd ABCd ABcd abcd Abcd aBCD ABCD
strict: Infeasible
relaxed: Infeasible
abcd Abcd aBcd ABcd abCd AbCd aBCd abcd Abcd aBcd ABCd ABcd abcd Abcd aBCD ABCD
{'A~~~': ['bcd', 'BCd'], 'a~~~': ['BCd', 'bcd']}
{'~B~~': ['acD', 'ACd'], '~b~~': ['ACd', 'acD']}

```

FIG. 9. Output of violation classification algorithm.

1. For each nonempty subset S of loci:
 - (a) For each assignment of alleles to these loci, collect into set R :
 - i. Filter the ordering of genotypes to the genotypes that contain the assigned alleles.
 - ii. Remove the loci in S from each genotype, yielding an ordering on genotypes over the remaining alleles.
 - (b) If set R contains more than one distinct ordering, a violation of directional consistency has been found; display it.

4. CONCLUSION

The transformation of traits to a natural scale can be performed effectively for either small or highly constrained systems. For genetic traits controlled by limited numbers of loci and alleles, our algorithm enumerates all possible trait structures and finds exact or error-minimizing linearizing transformations by formulating a constrained optimization program. We find that the fraction of possible distinct genetic traits satisfying simple criteria that can be fully or approximately linearized is high for small systems and falls with system complexity. In particular, multiallele diploid systems with few alleles require only directional consistency to be separable or approximately separable in most cases. Epistatic systems are generally more complex. While theoretical results in Sverdlov and Thompson (In Prep) suggest which biological traits are separable, computational methods can only be used to explain why a particular trait architecture cannot be transformed to a natural scale, but not to give broad statements about separability of classes of trait architectures.

We have demonstrated methods for classifying and solving, exactly and approximately, combinatorially defined trait architectures. Through linear programming, we recover the natural scale for an ordinal trait. Through our graph methods, we can generate trait architectures consistent with defined constraints, such as DCS.

These results demonstrate the feasibility of applying combinatorial methods to quantitative trait genetic systems and of classifying trait architectures into discrete categories, an approach which may help connect quantitative genetics to network models in systems biology. Further research in this direction would involve the enumeration and combinatorial classification of nonlinear features with various trait characteristics. This article appears to exhaust the combinatorial approach to diploidy and dominance with a small or moderate number of alleles, but computational methods can further classify epistatic systems, both separable/approximately separable and unseparable.

5. ACKNOWLEDGMENT

This research was supported in part by NIH grants R37 GM046255, T32 GM081062, and P01 GM099568.

6. AUTHOR DISCLOSURE STATEMENT

No competing financial interests exist.

REFERENCES

- Aho, A.V., Garey, M.R., and Ullman, J.D. 1972. The transitive reduction of a directed graph. *SIAM J. Comput.* 1, 131–137.
- Brightwell, G., and Winkler, P. 1991. Counting linear extensions. *Order*. 8, 225–242.
- Fusi, N., Lippert, C., Lawrence, N.D., et al. 2014. Warped linear mixed models for the genetic analysis of transformed phenotypes. *Nat. Commun.* 5, 4890.
- Heath, A., Martin, N., Eaves, L., et al. 1984. Evidence for polygenic epistatic interactions in man? *Genetics* 106, 719–727.
- Hill, W.G., Goddard, M.E., and Visscher, P.M. 2008. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet.* 4, e1000008.

- Kruskal, J.B. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29, 1–27.
- 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.
- Pruesse, G. and Ruskey, F. 1994. Generating linear extensions fast. *SIAM J. Comput.* 23, 373–386.
- Purdom, P. 1970. A transitive closure algorithm. *BIT Numer. Math.* 10, 76–94.
- Sverdlov, S. and Thompson, E.A. In Prep. The epistasis boundary: Linear vs. nonlinear genotype-phenotype relationships. *Theor. Popul. Biol.* (in preparation).
- Zuk, O., Hechter, E., Sunyaev, S.R., et al. 2012. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc. Natl. Acad. Sci.* 109, 1193–1198.

Address correspondence to:

*Serge Sverdlov
Department of Statistics
University of Washington
Box 354322
Seattle, WA 98195*

E-mail: serges@u.washington.edu