Statistical Physics Methods Provide the Exact Solution to a Long-Standing Problem of Genetics

Areejit Samal[1,2,3] and Olivier C. Martin[4,*]

[1] *Laboratoire de Physique Théorique et Modèles Statistiques (LPTMS), CNRS and Univ Paris-Sud, UMR 8626, F-91405 Orsay, France*
[2] *Max Planck Institute for Mathematics in the Sciences, 04103 Leipzig, Germany*
[3] *The Abdus Salam International Centre for Theoretical Physics, Trieste 34151, Italy*
[4] *INRA, UMR 0320/UMR 8120 Génétique Quantitative et Evolution–Le Moulon, F-91190 Gif-sur-Yvette, France*

(Received 13 November 2014; revised manuscript received 26 February 2015; published 9 June 2015)

Analytic and computational methods developed within statistical physics have found applications in numerous disciplines. In this Letter, we use such methods to solve a long-standing problem in *statistical genetics*. The problem, posed by Haldane and Waddington [Genetics 16, 357 (1931)], concerns so-called recombinant inbred lines (RILs) produced by repeated inbreeding. Haldane and Waddington derived the probabilities of RILs when considering two and three genes but the case of four or more genes has remained elusive. Our solution uses two probabilistic frameworks relatively unknown outside of physics: Glauber's formula and self-consistent equations of the Schwinger-Dyson type. Surprisingly, this combination of statistical formalisms unveils the *exact* probabilities of RILs for *any* number of genes. Extensions of the framework may have applications in population genetics and beyond.

DOI: 10.1103/PhysRevLett.114.238101 PACS numbers: 87.18.Wd, 02.50.Cw, 02.50.Sk, 05.40.-a

Statistical physics methods have fertilized numerous disciplines including complex networks [1], theoretical computer science [2], and Bayesian statistical inference [3]. They have also led to novel results in population genetics [4]. Here we use those methods to tackle an old problem of genetics involving recombinant inbred lines (RILs). A RIL is produced via repeated inbreeding of animals or plants until all genetic variability has been removed (see Fig. 1). The individuals produced in this way constitute a stable and permanently shareable genetic resource that is particularly useful for the identification of genes contributing to traits of interest [5]. These properties explain why production and exploitation of large populations of RILs have become major endeavors in the search for genetic determinants of diseases in mammals [6] and of agricultural traits in crops [7].

In this Letter, we consider plant RILs that are produced using *single seed descent* (SSD) which is an extreme form of inbreeding. One starts with two founding parents that are "homozygous" everywhere, i.e., for each pair of chromosomes, the two associated alleles are identical. This situation is schematically represented in Fig. 1 using the generation label F_0 and by displaying a single *pair* of chromosomes for each plant. The two parents being genetically different, their chromosomal contents are shown using different shadings. These two parents are then cross pollinated: one parent produces a female gamete while the other parent produces a male gamete. The fusion of the two gametes will lead to the *single* F_1 plant at the next generation. Consider going now from generation F_1 to generation F_2. Cross pollination is replaced by self-pollination: the single F_1 plant produces *both* the female gamete (g) and the male gamete (g'). This capability arises in almost all plants of agricultural interest. A subtlety now arises as shown in Fig. 1: a gamete can form a mosaic of the two chromosomes from which it is built. This phenomenon follows from the formation of "crossovers" between the two chromosomes during gamete formation. It can occur at all generations but in the case of going from F_0 to F_1 it simply has no visible effects. The process of producing a RIL is based on iterating the step when going from F_1 to F_2: *self-pollination* of a single F_n plant is used to produce a seed which develops into the single F_{n+1} plant, thus the term single seed descent. Note that once a chromosomal region has become homozygous (in the figure this corresponds to having locally the same shading for the two chromosomes) it stays so. (If a region is not homozygous, one says it is heterozygous.) Thus, because of chance, after enough generations, the plant becomes homozygous everywhere. The chromosomes of the resulting RIL are mosaics of the two parental chromosomes at F_0. Given many such RILs (cf. Ref. [8]), statistical inference can be used to identify the chromosomal regions responsible for parental differences in traits of interest [7].

Experimentally, one often determines a plant's genetic content at discrete positions or "loci"; we assign these an index i ranging from 1 to L (from left to right along the chromosome). Denote by a the allelic type (white) of the first parent and by A that (shaded) of the second parent. Then the *genotype* of parent a is $(a_1/a_1, a_2/a_2, \ldots, a_L/a_L)$ and that of parent A is $(A_1/A_1, A_2/A_2, \ldots, A_L/A_L)$ where the $\_i/\_i$ notation provides the allelic type on the two chromosomes for "locus" i. Figure 1 illustrates a case with $L = 3$ for which both gametes have crossovers when going

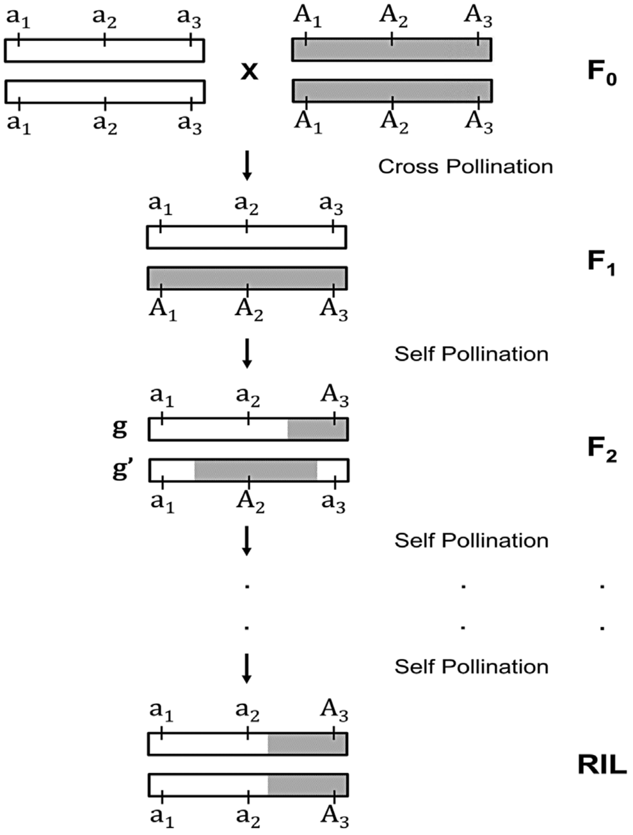


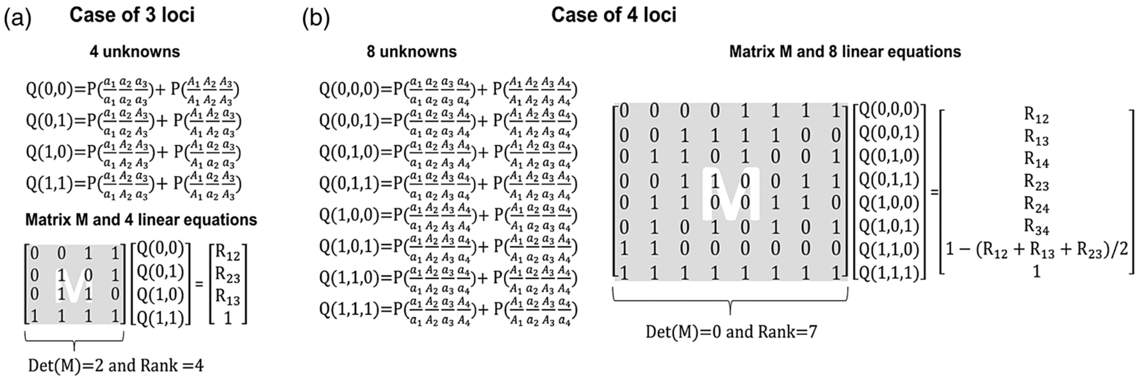
FIG. 1. Production of one recombinant inbred line. A chromosome pair is followed in each plant. A new generation results from two gametes that may mix genetic content as shown via g and g'. Tracking of allelic types (a and A) is displayed at three positions until no further change is possible.

from F_1 to F_2. The allelic type of the first gamete changes when going from locus two to locus three; one says that the interval (two, three) is "recombinant" or that there has been a recombination event between the two loci in that gamete.

In 1918 Robbins [9] determined the probabilities of two-locus RIL genotypes produced using SSD. Then, in 1931, Haldane and Waddington [10] simplified that derivation. Based on meiotic recombination rates independent of allelic content and of sex, they provided the celebrated Haldane-Waddington formula [10] giving the "RIL recombination rate" between two loci i and j, i.e., $R_{i,j} = 2r_{i,j}/(1+2r_{i,j})$ where $r_{i,j}$ is the (i,j) recombination rate *per meiosis*. Probabilities of all two-locus RIL genotypes are then directly obtained using the definition of the RIL recombination rate: $R_{i,j} = P(a_i/a_i, A_j/A_j) + P(A_i/A_i, a_j/a_j)$ which is the probability that the alleles will be recombined after enough inbreeding. By symmetry, $P(a_i/a_i, A_j/A_j) = P(A_i/A_i, a_j/a_j) = R_{i,j}/2$ and $P(a_i/a_i, a_j/a_j) = P(A_i/A_i, A_j/A_j) = (1-R_{i,j})/2$ ([11]).

In 1931 Haldane and Waddington [10] also showed that the two-locus RIL probabilities determine the ones for three loci. Over time, the results for two and three loci have been refined or extended to other kinds of crosses [12], but the case of four or more loci has proved to be inextricable. This fact appears as particularly puzzling since going from two to three loci is very simple and involves just standard algebra (see Fig. 2(a) and Ref. [13]). The point is that two- and three-locus RIL probabilities do *not* determine the four-locus probabilities (see Fig. 2(b) and Ref. [13]). Finding and exploiting this missing information has prevented researchers from extending the Haldane-Waddington result for over 80 years. In this Letter, we provide a solution to this challenge, deriving exact analytic formulas for the probabilities of RIL genotypes having *any* number of loci. The breakthrough is based on using two probabilistic frameworks borrowed from physics: the Schwinger-Dyson equations [14,15] and Glauber's formula [16].

Given that a RIL is homozygous at every locus, its genetic content can be specified in terms of a vector $\vec{S}$ of *spin* variables S_i, $i = 1, 2, \ldots, L$. Our convention, motivated by Ref. [17], is $S_i = 1$ if locus i is a_i/a_i and $S_i = -1$ if it is A_i/A_i. This notation is particularly convenient for writing the probability of any RIL genotype $\vec{S}$ in terms of averages of spin products. For example, if there is a single locus i, the probability that the spin has value s_i is $P(S_i = s_i) = E[(1 + s_i S_i)/2]$ where the average or expectation $E[\cdots]$ is taken over the distribution of the random variable S_i.



(a) Case of 3 loci
4 unknowns
$Q(0,0)=P(\frac{a_1\,a_2\,a_3}{a_1\,a_2\,a_3})+P(\frac{A_1\,A_2\,A_3}{A_1\,A_2\,A_3})$
$Q(0,1)=P(\frac{a_1\,a_2\,A_3}{a_1\,a_2\,A_3})+P(\frac{A_1\,A_2\,a_3}{A_1\,A_2\,a_3})$
$Q(1,0)=P(\frac{a_1\,A_2\,A_3}{a_1\,A_2\,A_3})+P(\frac{A_1\,a_2\,a_3}{A_1\,a_2\,a_3})$
$Q(1,1)=P(\frac{a_1\,A_2\,a_3}{a_1\,A_2\,a_3})+P(\frac{A_1\,a_2\,A_3}{A_1\,a_2\,A_3})$
Matrix M and 4 linear equations
$\begin{bmatrix}0&0&1&1\\0&1&0&1\\0&1&1&0\\1&1&1&1\end{bmatrix}\begin{bmatrix}Q(0,0)\\Q(0,1)\\Q(1,0)\\Q(1,1)\end{bmatrix}=\begin{bmatrix}R_{12}\\R_{23}\\R_{13}\\1\end{bmatrix}$
M
Det(M)=2 and Rank =4
(b) Case of 4 loci
8 unknowns
$Q(0,0,0)=P(\frac{a_1\,a_2\,a_3\,a_4}{a_1\,a_2\,a_3\,a_4})+P(\frac{A_1\,A_2\,A_3\,A_4}{A_1\,A_2\,A_3\,A_4})$
$Q(0,0,1)=P(\frac{a_1\,a_2\,a_3\,A_4}{a_1\,a_2\,a_3\,A_4})+P(\frac{A_1\,A_2\,A_3\,a_4}{A_1\,A_2\,A_3\,a_4})$
$Q(0,1,0)=P(\frac{a_1\,a_2\,A_3\,A_4}{a_1\,a_2\,A_3\,A_4})+P(\frac{A_1\,A_2\,a_3\,a_4}{A_1\,A_2\,a_3\,a_4})$
$Q(0,1,1)=P(\frac{a_1\,a_2\,A_3\,a_4}{a_1\,a_2\,A_3\,a_4})+P(\frac{A_1\,A_2\,a_3\,A_4}{A_1\,A_2\,a_3\,A_4})$
$Q(1,0,0)=P(\frac{a_1\,A_2\,A_3\,A_4}{a_1\,A_2\,A_3\,A_4})+P(\frac{A_1\,a_2\,a_3\,a_4}{A_1\,a_2\,a_3\,a_4})$
$Q(1,0,1)=P(\frac{a_1\,A_2\,A_3\,a_4}{a_1\,A_2\,A_3\,a_4})+P(\frac{A_1\,a_2\,a_3\,A_4}{A_1\,a_2\,a_3\,A_4})$
$Q(1,1,0)=P(\frac{a_1\,A_2\,a_3\,a_4}{a_1\,A_2\,a_3\,a_4})+P(\frac{A_1\,a_2\,A_3\,A_4}{A_1\,a_2\,A_3\,A_4})$
$Q(1,1,1)=P(\frac{a_1\,A_2\,a_3\,A_4}{a_1\,A_2\,a_3\,A_4})+P(\frac{A_1\,a_2\,A_3\,a_4}{A_1\,a_2\,A_3\,a_4})$
Matrix M and 8 linear equations
$\begin{bmatrix}0&0&0&0&1&1&1&1\\0&0&1&1&1&1&0&0\\0&1&1&0&1&0&0&1\\0&0&1&1&0&0&1&1\\0&1&1&0&0&1&1&0\\0&1&0&1&0&1&0&1\\1&1&0&0&0&0&0&0\\1&1&1&1&1&1&1&1\end{bmatrix}\begin{bmatrix}Q(0,0,0)\\Q(0,0,1)\\Q(0,1,0)\\Q(0,1,1)\\Q(1,0,0)\\Q(1,0,1)\\Q(1,1,0)\\Q(1,1,1)\end{bmatrix}=\begin{bmatrix}R_{12}\\R_{13}\\R_{14}\\R_{23}\\R_{24}\\R_{34}\\1-(R_{12}+R_{13}+R_{23})/2\\1\end{bmatrix}$
M
Det(M)=0 and Rank=7

FIG. 2. The two-and three-locus RIL probabilities do not completely specify the four-locus RIL probabilities. (a) The matrix of the linear equations relating the three-locus probabilities to two-locus probabilities (via the $R_{i,j}$'s) has rank 4. (b) A matrix giving linear equations relating the four-locus probabilities to two- and three-locus probabilities always has rank at most 7.

For L loci, the generalization of this formula, due to Glauber [16], is

$$P(\{S_1 = s_1, S_2 = s_2, \ldots, S_L = s_L\}) = E\left[\left(\frac{1 + s_1 S_1}{2}\right)\left(\frac{1 + s_2 S_2}{2}\right)\ldots\left(\frac{1 + s_L S_L}{2}\right)\right], \qquad (1)$$

where $E[\cdots]$ is the average over all possible RIL genotypes with their corresponding probabilities. Note that Eq. (1) is exact, the S_i need not be independent. The problem of finding the probabilities of all RIL genotypes is then solved if one can determine the expectation values of all spin products. When expanding the right-hand side of Eq. (1), expectation values of k-allelic products come with a sign equal to the product of the corresponding s_i values. For instance for $L = 4$, Eq. (1) leads to

$$P(\{S_1 = s_1, S_2 = s_2, S_3 = s_3, S_4 = s_4\}) = \frac{1}{16}\left(1 + \sum_{i<j} s_i s_j E[S_i S_j] + s_1 s_2 s_3 s_4 E[S_1 S_2 S_3 S_4]\right), \qquad (2)$$

where we have used the fact that the expectation of a product of an odd number of S_i's vanishes because of the global invariance $P(\vec{S}) = P(-\vec{S})$, corresponding to exchanging all a's and A's in RIL genotypes.

To explain our approach, we begin by solving the four-locus case ($L = 4$). Equation (2) shows that we need the expectations of 2- and 4-spin products. The 2-spin products are given by $E[S_i S_j] = 1 - 2R_{i,j}$ [11] so the only unknown is the 4-spin product $E[S_1 S_2 S_3 S_4]$ in direct correspondence with the situation described in Fig. 2(b). Our strategy to compute $E[S_1 S_2 S_3 S_4]$ is based on classifying the ways of going from the first generation of children (F_1) all the way to the RIL according to the genotype arising at the second generation of children (F_2) (Fig. 1). Performing this classification leads to

$$E[S_1 S_2 S_3 S_4] = \sum_g \sum_{g'} P(g) P(g') E_{g,g'}[S_1 S_2 S_3 S_4], \qquad (3)$$

where the sum is over all F_2 genotypes [each specified by the genotypes of its female (g) and male (g') gametes], $P(g)$ is the probability of producing a gamete of genotype g when going from F_1 to F_2, and $E_{g,g'}[S_1 S_2 S_3 S_4]$ is the expectation of the 4-spin product *when starting the inbreeding with an* F_2 *individual of genotype* (g, g'). Now the key point is that $E_{g,g'}[S_1 S_2 S_3 S_4]$ is equal to $E[S_1' S_2' S_3' S_4']$ when starting with the F_1 if one uses the following *substitution* rules for the S_i'. First, if locus i is homozygous in $G = (g, g')$ and has value s_i, then all descendants of G also have that value, so replace S_i' by s_i. Second, if locus i is heterozygous in G and is of the type a_i/A_i, the situation is the same as at F_1, so replace S_i' by S_i. Finally, if locus i is heterozygous in G and is of the type A_i/a_i, i.e., it is reversed compared to the F_1, replace S_i' by $-S_i$. These simple rules provide the way to relate expectations starting with an F_2 genotype to expectations starting with the F_1 genotype. The self-consistent Eq. (3) then becomes a Schwinger-Dyson (SD) equation [14,15] where the expectation value of the 4-spin product (on the left) is expressed (on the right) in terms of itself and of lower order spin-product averages. By summing the contributions of the 4^4 different F_2 genotypes in Eq. (3), we can extract the value for $E[S_1 S_2 S_3 S_4]$ and then our problem is solved, i.e., Eq. (2) provides all four-locus probabilities.

In Eq. (3) the sum over all F_2 genotypes involves the probabilities $P(g)$. If the crossovers arise independently as in Haldane's no interference model [18], then the summation in Eq. (3) can be performed by hand and very elegantly as follows. First we regroup the F_2 genotypes into classes according to which of their loci are heterozygous. For each class the associated contributions can be summed explicitly by mapping to a tree. To see how this works, consider for instance calculating the factor multiplying $E[S_1 S_2 S_3 S_4]$ on the right-hand side of the SD equation. This factor is obtained by considering the class of F_2 genotypes that are heterozygous (h) at all four loci and calculating the sum over those F_2 genotypes of the probability $[P(g)P(g')]$ times the sign from the substitution rule. Figure 3 represents the mapping of these genotypes, their probabilities and their signs onto a tree for the case where the first locus is of the type a_1/A_1. The loci are ordered from left to right

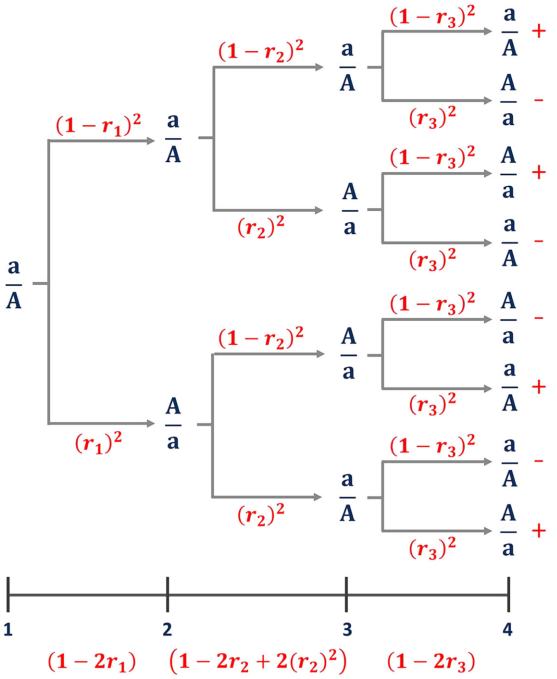


FIG. 3 (color online). Tree mapping of F_2 genotypes. F_2 genotypes map to paths from root to leaves of trees. The sign of a genotype is given on the right and its weight is the product of factors along the path. Summing over all paths of this tree leads to the factors shown at bottom.

and each F_2 genotype can be identified with a path from the left-most node to one of the right-most nodes (leaves of the tree). Because of the assumption of no crossover interference, recombination arises independently in each interval so that the probability of a genotype can be written as a product of factors, one for each interval. For any specified interval in Fig. 3, the two gametes are either both non-recombinant, leading to a factor $(1 - r_i)^2$, or both recombinant, leading to a factor r_i^2, where r_i is the recombination rate for a single meiosis in the interval $(i, i + 1)$. The probability of a F_2 genotype is then given by the product of such factors along the path as displayed in Fig. 3, times $1/4$ coming from the probability that the first locus is of the type a_1/A_1. Adding the contributions of all genotypes of the tree shown in Fig. 3 can be done by recurrence [19]. Using the fact that the tree rooted at A_1/a_1 gives rise to the same calculation as for Fig. 3, one concludes that the class of heterozygous genotypes on the right-hand side of Eq. (3) contributes a total of $(1 - 2r_1)[(1 - r_2)^2 + r_2^2](1 - 2r_3)/2$ times $E[S_1S_2S_3S_4]$.

The other classes can be treated by the same mapping technique. Consider for instance the class of F_2 genotypes homozygous at all loci. It is easy to see that it leads to exactly the same result as the class just treated except that $E[S_1S_2S_3S_4]$ is replaced by 1 [19]. Going on to the classes which are mixed (with both homozygous and heterozygous loci), only those having two adjacent loci homozygous and two adjacent loci heterozygous lead to nonzero contributions [19]. In those cases, between the second and third locus, there is one and only one recombinant gamete, whereas in the previous calculation in that interval g and g' were both recombinant or both non-recombinant. Thus the previously derived term $[(1 - r_2)^2 + r_2^2]$ has to be replaced by $2r_2(1 - r_2)$ here (Fig. S2 in Ref. [19]). Collecting the results from all classes of F_2 genotypes leads to the four-locus SD equation:

$$\begin{aligned} E[S_1S_2S_3S_4] = {} & \frac{(1-2r_1)((1-r_2)^2 + r_2^2)(1-2r_3)}{2} \\ & \times (E[S_1S_2S_3S_4] + 1) \\ & + \frac{(1-2r_1)(2(1-r_2)r_2)(1-2r_3)}{2} \\ & \times (E[S_1S_2] + E[S_3S_4]). \end{aligned} \tag{4}$$

Although the expectation of the 4-spin product arises on both sides of this equation, extracting this quantity in terms of the averages of 2-spin products is straightforward. In summary, from Eq. (2), using Eq. (4) and the formula $E[S_iS_j] = 1 - 2R_{i,j}$, one obtains the long-searched-for exact analytic expressions for four-locus RIL genotype probabilities.

The overall framework, including the mapping of F_2 genotypes to trees, extends to any number of loci. For five loci, no new SD equation is needed since the expectation $E[S_1S_2 \ldots S_L]$ vanishes when L is odd. For six or seven loci, Eq. (1) shows that we need expectations of 2-, 4- and 6-spin products. We have determined the 2- and 4-spin products above, and the mapping onto trees for computing the 6-spin product follows exactly the same logic as for the 4-spin product [20]. More generally, when going from L to $L + 2$ loci, the only new unknown is the expectation of the product of all spins. Interestingly, the SD equations follow simple patterns [21]. Based on these patterns, we have written a computer program that takes as input the list of genetic positions of L loci and computes the probability of all L-locus RIL genotypes [22]. Lastly, the approach is easily extended to the case where male and female recombination rates differ [23].

These exact rather than approximate probabilities of multilocus genotypes could be used in a number of situations in which RIL probabilities are needed. For instance when building genetic maps, the ordering of markers relies on comparing likelihoods of multilocus genotypes, generally approximated by products of pairwise recombination rates over putatively adjacent loci [24]. The same approximation is routinely applied in algorithms for detection of quantitative trait loci using interval or composite interval mapping [25]. Similarly, when genotypes or haplotypes must be inferred or imputed because of missing information or because markers are not sufficiently dense [26], determining the most likely assignment requires comparing multilocus genotype probabilities. Moving beyond RILs, it is possible that our framework will unveil ways to perform calculations of multilocus probabilities in more general population genetics contexts [27] where the main difficulty comes from having a potentially infinite number of generations. That situation arises when one is interested in fixation probabilities, steady-state multilocus frequencies, or distribution times of the most recent common ancestor [28–30].

In 1931 Haldane and Waddington [10] provided the exact two-locus probabilities for successive generations (F_2, F_3, …) based on recursion formulas from which they were able to extrapolate to RILs, i.e., to an infinite number of generations. In the present work, we have instead directly treated the RIL situation, exploiting Eq. (1) due to Glauber [16] and self-consistent equations of the Schwinger-Dyson type [14,15]. *A posteriori*, it is quite surprising that these mathematical tools had not been used before to generalize the Haldane-Waddington formula. Perhaps just as surprising is their remarkable efficiency for solving this long-outstanding problem.

We thank D. de Vienne, C. Dillmann, M. Gromov, F. Hospital, S. Majumdar, N. Morozova, and B. Servin for comments and references. AS and OCM acknowledge financial support from CNRS GDRE 513, ICTP, and INRA.

*olivier.martin@moulon.inra.fr

[1] R. Albert and A.-L. Barabasi, Rev. Mod. Phys. **74**, 47 (2002).

[2] M. Mezard, G. Parisi, and R. Zecchina, Science **297**, 812 (2002).

[3] U. von Toussaint, Rev. Mod. Phys. **83**, 943 (2011).

[4] V. Mustonen and M. Lassig, Proc. Natl. Acad. Sci. U.S.A. **107**, 4248 (2010); R. A. Neher and B. I. Shraiman, Rev. Mod. Phys. **83**, 1283 (2011); E. Brunet and B. Derrida, J. Stat. Mech. (2013) P01006.

[5] J. F. Crow, Genetics **176**, 729 (2007).

[6] Complex Trait Consortium: http://www.complextrait.org/.

[7] E. S. Buckler *et al.*, Science **325**, 714 (2009).

[8] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.114.238101 for the section *Production of a recombinant inbred line via single seed descent* and Fig. S1.

[9] R. B. Robbins, Genetics **3**, 375 (1918).

[10] J. B. S. Haldane and C. H. Waddington, Genetics **16**, 357 (1931).

[11] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.114.238101 for the section *Rederiving the Haldane-Waddington formula via the new framework and the case of sex-specific recombination rates*.

[12] C. R. Winkler, N. M. Jensen, M. Cooper, D. W. Podlich, and O. S. Smith, Genetics **164**, 741 (2003); K. W. Broman, Genetics **169**, 1133 (2005); O. C. Martin and F. Hospital, Genetics **173**, 451 (2006).

[13] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.114.238101 for the section *Probabilities for 2 and 3 loci do not determine those for 4 loci*.

[14] F. Dyson, Phys. Rev. **75**, 1736 (1949).

[15] J. Schwinger, Proc. Natl. Acad. Sci. U.S.A. **37**, 452 (1951).

[16] R. J. Glauber, J. Math. Phys. (N.Y.) **4**, 294 (1963).

[17] M. Slatkin, Genetics **72**, 157 (1972).

[18] J. B. S. Haldane, Journal of Genetics **8**, 299 (1919).

[19] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.114.238101 for the section *Mapping all F2 genotypes with 4 loci to trees* and Figs. S2, S3.

[20] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.114.238101 for the section *The Schwinger-Dyson equations for 6 loci and beyond* and Figs. S4–S7.

[21] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.114.238101 for the section *Useful properties for simplifying the derivation of the Schwinger-Dyson equations*.

[22] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.114.238101 for a computer program in C to compute the probabilities of RIL genotypes for any number of loci. Since the code determines all 2^L expectations of k-spin products, $(0 \leq k \leq L)$, it runs in a time growing exponentially with L. Empirically, the time is multiplied by 10 when L increases by 2, but even at $L = 14$ running time is less than a second.

[23] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.114.238101 for the section *Generalizing the formulas to sex-specific recombination rates*.

[24] E. S. Lander, P. Green, J. Abrahamson, A. Barlow, M. Daly, S. Lincoln, and L. Newburg, Genomics **1**, 174 (1987).

[25] Z. B. Zeng, Genetics **136**, 1457 (1994); S. Sen and G. A. Churchill, Genetics **159**, 371 (2001).

[26] B. Servin and M. Stephens, PLoS Genetics **3**, e114 (2007).

[27] F. Zanini and R. A. Neher, Bioinformatics **28**, 3332 (2012).

[28] J. F. C. Kingman, Stoch. Processes Appl. **13**, 235 (1982).

[29] B. Derrida and B. Jung-Muller, J. Stat. Phys. **94**, 277 (1999).

[30] K. Lohse, R. J. Harrison, and N. H. Barton, Genetics **189**, 977 (2011).