Competition between recombination and epistasis can cause a transition from allele to genotype selection

Richard A. Neher[a] and Boris I. Shraiman[a,b,1]

[a]Kavli Institute for Theoretical Physics, and [b]Department of Physics, University of California, Santa Barbara, CA 93106

Edited by Curtis G. Callan, Jr., Princeton University, Princeton, NJ, and approved February 27, 2009 (received for review December 14, 2008)

Biochemical and regulatory interactions central to biological networks are expected to cause extensive genetic interactions or epistasis affecting the heritability of complex traits and the distribution of genotypes in populations. However, the inference of epistasis from the observed phenotype–genotype correlation is impeded by statistical difficulties, while the theoretical understanding of the effects of epistasis remains limited, in turn limiting our ability to interpret data. Of particular interest is the biologically relevant situation of numerous interacting genetic loci with small individual contributions to fitness. Here, we present a computational model of selection dynamics involving many epistatic loci in a recombining population. We demonstrate that a large number of polymorphic interacting loci can, despite frequent recombination, exhibit cooperative behavior that locks alleles into favorable genotypes leading to a population consisting of a set of competing clones. When the recombination rate exceeds a certain critical value that depends on the strength of epistasis, this ''genotype selection'' regime disappears in an abrupt transition, giving way to ''allele selection''—the regime where different loci are only weakly correlated as expected in sexually reproducing populations. We show that large populations attain highest fitness at a recombination rate just below critical. Clustering of interacting sets of genes on a chromosome leads to the emergence of an intermediate regime, where blocks of cooperating alleles lock into genetic modules. These haplotype blocks disappear in a second transition to pure allele selection. Our results demonstrate that the collective effect of many weak epistatic interactions can have dramatic effects on the population structure.

gene interactions | population genetics

Selection acting on genetic polymorphisms in populations is a major force of evolution (1–4) and it is possible to identify specific loci under positive selection, e.g., the *Adh* locus in *Drosophila* (4). However, the attribution of fitness differentials to specific allelic variants and combinations remains a great challenge (5). Efforts to correlate quantitative phenotypes with genetic polymorphisms typically identify a small number of loci with a significant contribution to the observed phenotypic variance, but leave much of the variance unaccounted for (6). This unaccounted variance is believed to arise from a large number of loci with small individual contributions, or be due to epistasis and quite likely involves both effects. New studies accumulate evidence that epistasis is widespread and accounts for a significant fraction of phenotypic variation, e.g., in yeast (7–9). Additional evidence for epistasis comes from crosses of mildly diverged strains, where the recombinant progeny often has reduced average fitness, i.e., displays outbreeding depression. The reduction in fitness is attributed to the breakdown of favorable combinations of alleles in the ancestral strains (10). Outbreeding depression is often observed in partly selfing organisms such as *Caenorhabditis elegans* (11) or plants (12), species with strong geographic isolation such copepod (13) or facultatively mating organisms such as yeast (14). Although most recombinant genotypes are less fit, novel genotypes that perform better than either parental strain can be generated as well (15). Such outcrossing events could play an important role in evolution.

Competition between epistatic selection and recombination, explicit in the outbreeding depression phenomenon, is the focus of the present study. In the presence of epistasis, selection, by increasing the frequency of favorable genotypes, establishes correlations between alleles at different loci. Recombination however reshuffles alleles and randomizes genotypes breaking up coadapted loci. Because the recombination rate between any 2 loci is largely determined by their physical distance on the chromosome, the effect of genetic interactions depends on gene location. It is known that functionally related genes tend to cluster (16, 17), suggesting selection on gene order. Furthermore, chromosomes have regions of infrequent recombination, interspersed with recombination hotspots (18). Does selection have a hand in defining low recombination regions? To understand how evolution shaped genomes as we observe them today, we have to tackle the problem of how selection acts on many interacting polymorphisms for a large range of recombination rates (19).

Standing variation harbored in natural population provides important raw material for selection to act upon, in particular after a sudden change in environments or hybridization events (20). In such a situation, selection will reduce genetic variation until a new mutation-selection equilibrium is reached. Here, we show that the selection dynamics on standing variation at a large number of loci can be strongly affected by epistasis, even if the individual contribution of each locus is small. The competition between selection on epistasis and recombination gives rise to 2 distinct regimes at high and low recombination rates separated by a sharp transition. The population dynamics in the two regimes is illustrated in Fig. 1 *A* and *B*: (*i*) the "clonal competition" (CC) regime, which occurs for recombination rates $r < r_c$ and (*ii*) the quasi linkage equilibrium (QLE) regime for $r > r_c$. The different nature of the two regimes is best understood by considering the limiting cases of no and frequent recombination. In the case of purely asexual reproduction, selection operates on entire genotypes and results in clonal expansion of the fitter ones. The genetic variation present in the initial population is lost on a timescale inversely proportional to the average magnitude of fitness differentials between genotypes present in the population. Successful genotypes persist in time, which is apparent as continuous broad stripes of one color in Fig. 1*A*. The amplification of a small number of fit genotypes induces strong correlations or linkage disequilibrium among loci. In presence of epistasis, a little recombination does not change this picture qualitatively, because most recombinant genotypes are less fit than the prevailing clones and novel successful clones are rare. Never-

Author contributions: R.A.N. and B.I.S. designed research, performed research, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: shraiman@kitp.ucsb.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0812560106/DCSupplemental.

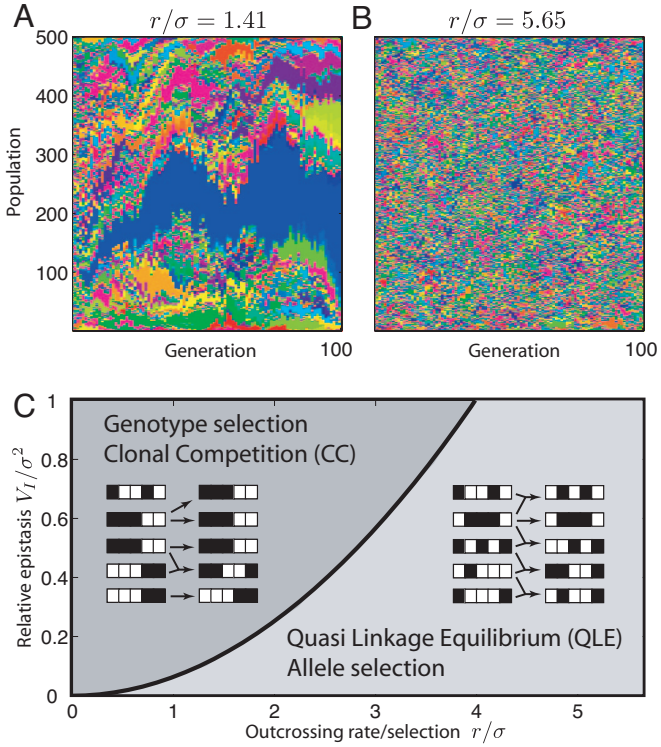

Fig. 1. The two regimes of sexual reproduction. (*A* and *B*) The simulated time course of the genotype distribution in a population of 500 individuals with epistatic fitness variance $V_I = \sigma^2 = 0.005$ and outcrossing rate $r = 0.1$ (*A*) and $r = 0.4$ (*B*; RE model defined below). Like genotypes are assigned the same color and stacked on top of each other. (*C Insets*) Sketches illustrating the population dynamics in the 2 cases. At low outcrossing rates, fit genotypes can proliferate. The genotype distribution rapidly coarsens and clones form (horizontal stripes in *A*). With frequent outcrossing, genes are rapidly reshuffled and genotypes do not persist over many generations, resulting in the pointillist pattern in *B*. Fixation happens at later time and is not shown. (*C*) The two regimes are separated by a sharp boundary set by the strength of epistasis. For $r < r_c$, the population dynamics is described by clonal competition (CC); for $r > r_c$ by quasi linkage equilibrium (QLE).

theless recombination is very important because it continuously introduces new genotypes leading to an increase in fitness attained by the population at long times. In the limit of high recombination genotypes are short-lived and essentially unique, resulting in a "pointillist" color pattern in Fig. 1*B*. Each allelic variant is therefore selected on the basis of its effect on fitness, averaged over many possible genetic backgrounds. The time scale on which allele frequencies change is given by the inverse of these marginal fitness effects. The term "linkage equilibrium" in QLE refers to the negligible correlations between loci, which are constantly reshuffled by recombination.

As we show below, the transition between the two regimes sharpens as the number of segregating loci L increases. The sharpening of the transition is related to the different scaling of the time scale of selection in the two regimes. For large L, the marginal fitness effects of individual loci become small compared with fitness differentials among individuals (assuming they are all of similar size, this ratio decreases as $\sim 1/\sqrt{L}$). Hence, the dynamics in the QLE regime slows down compared with the CC regime as L increases. The CC and QLE regimes correspond to different regions of the parameters space spanned by the relative strength of epistasis and the ratio of outcrossing or recombination rate to the strength of selection, as sketched in Fig. 1*C*. The QLE dynamics was first described by Kimura (21) in the limit of weak selection/fast recombination for a pair of biallelic loci and subsequently generalized to multiloci systems (22, 23). The possibility of a collective behavior involving linkage disequilibrium among many loci and selection effectively acting on the whole chromosome as a unit has been pointed out before in the context of overdominance by Franklin and Lewontin (24) in the strong selection limit. However, these studies of the two different limits do not reveal the breakdown of QLE and the transition to CC as the generic behavior of multilocus epistatic systems.

To underscore the general nature of the results, have studied 2 different models of epistasis. The first model follows the common treatment of epistasis in quantitative traits, which assumes that the epistatic contribution to fitness is disrupted when the parental genes are mixed in sexual reproduction (25, 26). This assumption becomes exact when the epistatic component of fitness of a specific genotype is a random number (which depends on the genotype, but is fixed in time) and we call this model the random epistasis (RE) model. Within the RE model, any change in the genotype randomizes the epistatic component of fitness so that the latter is not heritable when nonidentical parents mate. It is, however, faithfully passed on to the offspring in asexual reproduction. For the RE model, genomes are propagated asexually with probability $1 - r$ and with probability r are a product of mating where all genes are reassorted, as would be exactly correct if all genes were on different chromosomes. This model of facultative mating approximates reproductive strategies common in fungi (e.g., yeast) or nematodes and plants. As a more realistic alternative, we also study a model with only pairwise interactions between loci (27). This pairwise epistasis (PE) model allows epistatic contribution to be partly heritable, because interacting pairs have a chance to be inherited together (28). For the PE model, we assume that all genes are arranged on a single chromosome with a uniform cross-over rate ρ, which allows us to explore haplotype block formation and implications for recombination rate evolution.

The strength of selection is determined by the variance σ^2 of the distribution of fitness in the population. Within our models, the fitness $F(g)$ of a genotype g is the sum of an additive component $A(g)$ representing independent contributions of alleles and an epistatic part $E(g)$. For the RE model, the latter is a random number drawn from Gaussian distribution, whereas for the PE model it is a sum of pairwise interactions with random coefficients f_{ij}. The variances V_A and V_I of the distributions of $A(g)$ and $E(g)$ add up to σ^2 and their relative magnitude determines the importance of additive effects compared with epistasis. The two different models and their parameters are given explicitly in *Methods*. For the sake of simplicity, we assume haploid genomes. Random and pairwise epistasis represent 2 opposite extremes in the complexity of epistasis. Although the pairwise model is more realistic, the generic behavior is most clearly demonstrated using the RE model with random gene reassortment and facultative mating.

Results

Two Regimes of Selection Dynamics. We performed extensive computer simulation of our two models for different relative strength of epistasis, $L = 25$–200 loci and populations sizes between $N = 500$ and 10^6. We initialize simulations in a genetically diverse state as would result from multiple crossings of 2 diverged strains and examine the evolution under selection and recombination. The two regimes differ strongly in the amount of linkage disequilibrium (LD) (see *Methods*) build up by selection. Fig. 2*A* shows the average LD per locus pair for the RE model as a function of the outcrossing rate r. For $r < r_c$, the LD per locus pair is of order 1 and independent of L or N, indicating genome-wide LD. LD builds up despite a large number of different genotypes in the population interbreeding constantly. For $r > r_c$, the LD is much smaller, with the observed value determined by the sampling noise due to the finite population size (see Fig. 2*A Inset* and Fig. S1). Similar behavior occurs in the PE model, as shown in Fig. 2*B*. Above a critical recombination rate ρ_c, the observed linkage disequilibrium is time independent and

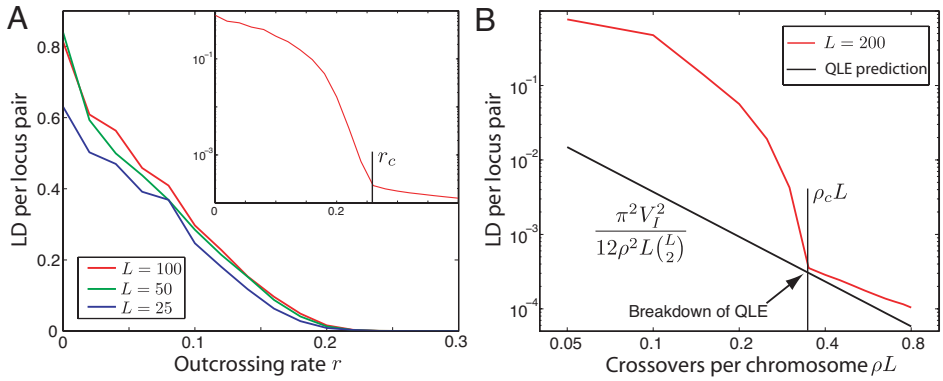

Fig. 2. The clonal competition regime is characterized by extensive linkage disequilibrium. (*A*) Random epistasis model: For small *r*, the LD per locus pair is of order 1 and fairly independent of *L*. (*Inset*) Data for $L = 100$ on a logarithmic scale and a mark at the value of r_c. The LD for $r > r_c$ is due to sampling noise, see Fig. S1. (*B*) Pairwise epistasis model. For pairwise epistasis, the QLE approximation gives explicit predictions for LD, which describes the observed LD very accurately for $\rho > \rho_c$, black line. For $\rho < \rho_c$, LD is a much larger than the QLE prediction. For *A* and *B*, LD is measured when allelic entropy has decayed 30% from the initial value ($\sigma^2 = 0.005$, $V_A = 0.1\sigma^2$ and $V_I = 0.9\sigma^2$). In *A*, $N = 10^5$, and the data shown are averaged over 100 realizations. To avoid boundary and finite size effects, we used $N = 10^6$, assumed a circular chromosome for *B*, and averaged over 10 realizations.

well described by the QLE approximation (21, 22) (straight line) (see *SI Appendix*). The QLE approximation (in the high ρ/σ limit) predicts LD to be proportional to the strength of pairwise epistasis. Below ρ_c, the observed LD is dramatically larger than the QLE expectation. Here, recombination is sufficiently infrequent such that genotypes with a synergistic alleles are amplified faster than they are taken apart by recombination, see below. As a result, the few fittest genotypes grow exponentially in number, leading to the strong correlation in the occurrence of cooperating alleles, independent of physical linkage (i.e., proximity on the chromosome). This extensive LD leads to a complete failure when extrapolating results valid in the high recombination regime across the transition. The relevant quantity that determines whether fit genotypes can be maintained is the probability that no cross-over occurs, which is given by $e^{-\rho L}$. Hence, ρ_c is inversely proportional to L.

Self-Consistency Condition for QLE. The fitness of a genotype can be decomposed as $F = A + E$, where A is the heritable additive part and E is the nonheritable epistatic part. As a coarse-grained descriptor of the population, we consider the joint distribution $P(A, E; t)$ of the fitness components. In the QLE state, $P(A, E; t)$ evolves approximately as

$$\partial_t P(A, E; t) = (F - \bar{F} - r)P(A, E; t) + r\rho(E)\,\vartheta(A; t) \qquad [1]$$

The first term accounts for the exponential growth of genotypes with fitness advantage $F - \bar{F}$ and the loss due to recombination at rate *r*. The second term accounts for the production of genotypes through recombination. To a good approximation, the distribution of A among recombinant offspring is identical to that among the parents $\vartheta(A) = \int dE\ P(A, E)$, which in turn is approximately Gaussian (29). The distribution of E among recombinant offspring is independent of the parents and a random sample from the distribution of epistatic fitness $\rho(E)$, which in our models is a zero-centered Gaussian. The latter is exactly true for the RE model and holds approximately for the PE model, where the correlation of E between ancestor and offspring halves every generation (28).

Eq. **1** admits the factorized solution $P(A, E; t) = \vartheta(A; t)\omega(E)$ with $\partial_t\,\vartheta(A; t) = (A - \bar{A})\vartheta(A; t)$ and a time-independent distribution of E

$$\omega(E) = \frac{r\rho(E)}{r + \bar{E} - E}, \qquad [2]$$

where $\bar{E}$ is determined by the condition that $\omega(E)$ has to be normalized. This solution exists only if $E < r + \bar{E}$ for all genotypes; otherwise, fit genotypes escape recombination and grow as clones. These two scenarios are illustrated in Fig. 3.

The normalization condition can be fulfilled only if *r* is larger than some r_c. Note that $\rho(E)$ has to go to 0 faster than linear for r_c to exist. The value of r_c is proportional to the maximal E and hence proportional to the strength of epistasis $\sqrt{V_I}$. However, it is not the absolute maximum of E among all possible 2^L genotypes that determines r_c, but the maximal E that is encountered by the population before fixation. Hence, r_c depends on the population size and the functional form of this dependence is determined by the upper tail of the distribution $\rho(E)$. For the Gaussian distribution used here, $r_c \approx \sqrt{2V_I \ln(rN\tau)}$, where τ is the time scale of QLE dynamics discussed below. The product $rN\tau$ then is the number of genotypes generated through recombination before fixation. A more detailed discussion is given in the *SI Appendix*.

The breakdown of the QLE state has some similarity to the error-threshold transition of a quasi-species model (30) in a rugged fitness landscape (31): Recombination of epistatic loci acts as deleterious mutations and prevents the emergence of quasi-species or clones (32, 33) for $r > r_c$.

Maintenance of Genetic Diversity. The transition between the two regimes leaves its imprint in virtually every quantity of interest in population genetics. For instance, the characteristic time for the decay of genetic diversity, τ (which we quantify via allele entropy, see *Methods*) scales differently with L in the two regimes, as shown in Fig. 4*A*. At low outcrossing rates, τ depends only on the total variance in fitness and neither on the number of loci nor the relative strength of additive contributions. This is consistent with the notion that in the CC regime genotypes are the units on which selection acts. With more frequent outcrossing, τ tends to be larger for weak additive contributions and large L. Beyond a certain outcrossing rate r_c, τ becomes independent of r attaining a value inversely proportional to the additive contribution of the individual loci independent of V_I (black diamonds in Fig. 3*A*). This observation confirms our assertion that for $r > r_c$, outcrossing decouples the loci and that the allele frequencies evolve independently under the action of the additive component of fitness. Given an additive variance V_A, the typical single locus fitness differential is $f \sim \sqrt{V_A/L}$ such that τ grows as $\sqrt{L}$ for $r > r_c$. To uncover the universal behavior in the vicinity of the transition in the limit of large genomes, we show that the data for different V_I, V_A, and L collapses

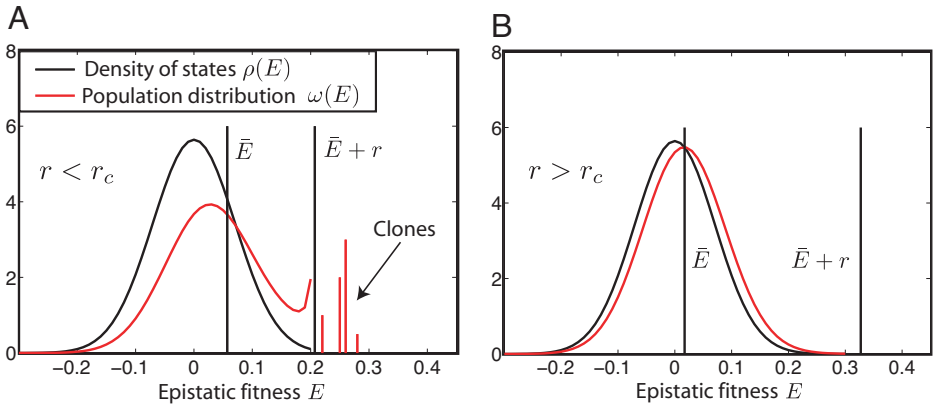

Fig. 3. The break-down of QLE. (*A*) When the recombination rate decreases below r_c, some individuals have epistatic fitness E larger than $\bar{E}+r$, and the QLE solution for the distribution of epistatic fitness in the population breaks down. Individuals to the right of $\bar{E}+r$ form clones that grow exponentially and the population condenses into a small number of genotypes. (*B*) For $r > r_c$, even the largest epistatic fitness contributions do not result in a growth advantage that exceeds the recombination rate.

onto a single master curve after appropriate rescaling of the axis, see Fig. S2. This scaling collapse demonstrates the existence of a sharp transition in the limit $L \to \infty$, the scaling of τ with $\sqrt{L}$ and shows that r_c is proportional to $\sqrt{V_I}$, as expected from the self-consistency argument outlined above and sketched in Fig. 1*C*. The suppression of allele dynamics by $1/\sqrt{L}$ in the QLE regime is at the basis of Fisher's infinitesimal model put forward to explain sustained response to selection (6). In one generation, the allele frequencies change by approximately f, which can be sustained over $\sim f^{-1}$ generations. The mean fitness increases by V_A per generation, consistent with Fisher's theorem (23, 26). Our results show, that epistasis causes the breakdown of the infinitesimal model for $r < r_c$. The pairwise epistasis model is more complex than the random epistasis model, because the partition of the fitness variance in additive and epistatic contribution depends on the allele frequencies and epistasis is "converted" into additive fitness as the population approaches fixation (34).

The properties of the genotype that will eventually fixate in the population depend on the regime in which it was obtained. We find, that the fitness of this fixated genotype depends nonmonotonically on the outcrossing rate and peaks just below the transition, see Fig. 4*B*. This can be understood as follows. Without recombination, the final state can be no fitter than the fittest genotype initially present. With some recombination, the population explores a greater number genotypes, potentially finding ones with higher fitness so that the fitness of final state increases with r in the CC regime. A similar benefit of infrequent recombination due to exploration of genotype space has been studied in the context of virus evolution for additive fitness functions (35). As genotype selection gives way to allele selection, different loci decouple and the epistatic contribution to fitness is missed, leading to possible fixation of less fit genotypes and a sharp drop of the final fitness as r approaches r_c. The dependence of the final fitness on the population size N highlight the distinct properties the dynamics in the two regimes: In the QLE regime, the final fitness is virtually identical for different N. This is a consequence of the fact that the relevant dynamical variables are allele frequencies, which are well sampled by $\mathcal{O}(N)$ individuals. Fluctuations of the allele frequencies are therefore negligible and the dynamics is essentially deterministic. This is different in the CC regime, where the dynamics is driven by the generation of rare, exceptionally fit genotypes. The rate, at which genotypes are generated is proportional to the N, resulting in a pronounced dependence on the population size. QLE ceases to be deterministic once the marginal fitness effects become comparable to inverse population size and random genetic drifts overwhelms selection (see Fig. S3).

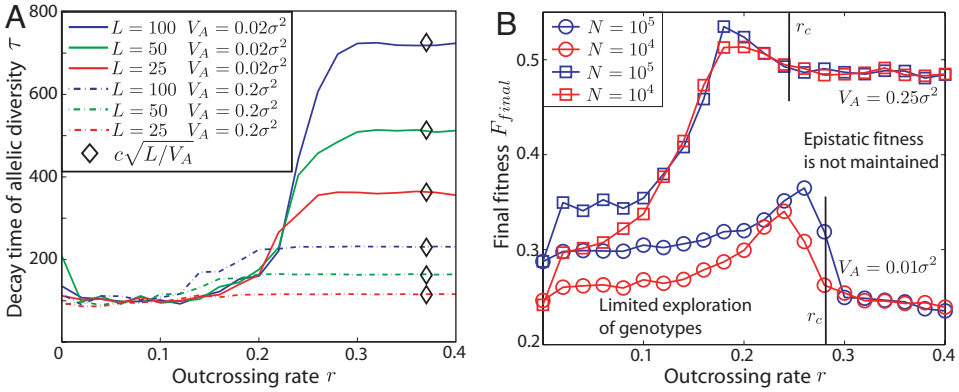

Fig. 4. Decay of genetic diversity and success of selection for RE model. (*A*) The time τ it takes to reduce the allelic entropy by 30% as a function of r for different parameters. For small r, τ is independent of L but increases with r. For $r > r_c$, τ settles at $c\sqrt{L/V_A}$ (black diamonds) in accord with the theoretical prediction for single locus dynamics (with c a fitting parameter). Additional data for $V_A = 0$, $V_A = 0.5\sigma^2$ and a collapse confirming the scaling of τ is shown in Fig. S2. (*B*) The fitness of the fixated genotype F_{final} as a function of r for 2 different strength of epistasis. At $r = 0$, the final fitness only depends on the population size N and is independent of the strength of epistasis. F_{final} increases with r, followed by a pronounced drop right below r_c. Above r_c, F_{final} is almost constant and is independent of N. In both *A* and *B*, $\sigma^2 = 0.005$. Data are averaged over 25 realizations in *A* and over 100 realizations in *B*. L = 100 in *B*.

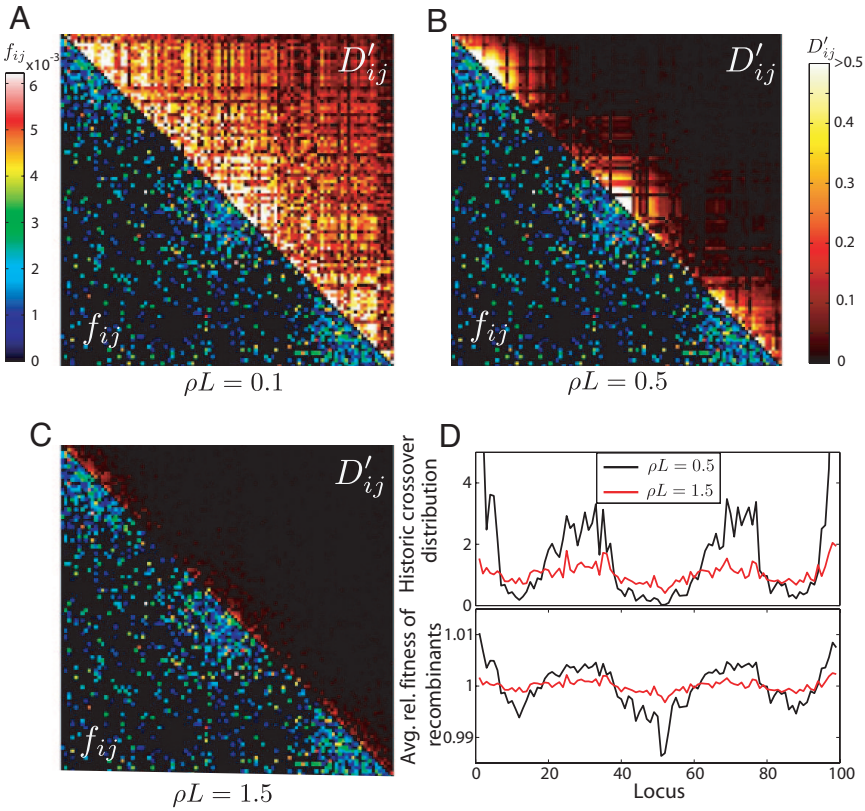

Fig. 5. Clonal competition, modular selection and quasi linkage equilibrium. (*A–C*) LD, measured as D'_{ij} between 2 loci *i* and *j*, is shown above the diagonal for a linear chromosome of length $L = 100$ at 3 different cross-over rates ρ. The interaction matrix f_{ij} is shown below the diagonal. At low ρ (*A*), the sparse long range interactions suffice to produce genome wide LD. At intermediate ρ (*B*), distant parts of the genome are decoupled, but the more strongly interacting clusters still show high LD, which vanishes at even higher recombination rates (*C*). (*D*) (*Upper*) The distribution of historic cross-overs. (*Lower*) The relative fitness of recombinants as a function of the cross-over location. LD was measured when allelic entropy was at 90% of the initial value, $\sigma^2 = V_I = 0.005$ and $N = 10^6$.

Selection on Genetic Modules. So far, we assumed that each pair of loci is equally likely to interact epistatically, regardless of their physical distance on the chromosome. However, there is evidence that the order of genes along the chromosome is far from random and that related genes tend to cluster (16, 17). To emulate such a situation we use the PE model and construct an interaction matrix f_{ij} where arbitrary pairs interact with a small probability while clusters of neighboring genes interact with a high probability (see *Methods*). For such a hierarchical epistatic structure, we observe, as a function of increasing cross-over rate ρ, a sequence of 2 transitions that define, sandwiched between CC and QLE, an intermediate Modular Selection (MS) regime, where the genome-wide LD characteristic of the CC regime has broken down to a set of modular blocks that are in quasi linkage equilibrium with each other. The resulting linkage disequilibrium patterns are shown in Fig. 5. The observed block structure of LD in the MS regime resembles haplotype blocks (18, 19), which are normally associated with regions of little recombination flanked by recombination hotspots. Indeed, the cumulative recombination history of the chromosomes in the population show a very heterogenous recombination distribution, as shown in Fig. 5*D*. However, here the origin of these blocks is not intrinsically low recombination (i.e., physical linkage) but the collective effect of epistatic selection: The surviving individuals have recombined more often in regions of low epistasis than in regions of high epistasis, even though the attempted cross-overs are uniformly distributed along the chromosome. Clusters of epistatic interaction can therefore exert selective pressure to lower recombination within the cluster. This lack of recombinant survival has been observed in experiments with mice (36), where inbreeding results in strong selective pressure on localized clusters of genes generating blocks with high LD and reduced effective recombination.

Conclusion

We have shown that the competition of epistatic selection and recombination can give rise to distinct regimes of population dynamics, separated by a transition that becomes sharp for large number of interacting loci. The QLE and CC regimes are realizations of the opposing views on evolution of R. A. Fisher and S. Wright. For $r > r_c$ alleles are selected for the their additive contributions while selection acts on whole genotypes for $r < r_c$. The fundamental differences between these two regimes show up most clearly in the different scaling properties of the total LD and the decay time of genetic diversity. In the low recombination regime, LD is produced independent of physical linkage by the collective effect of many interactions. In the high recombination regime, LD can be attributed to specific interactions between pairs of loci and its value, determined by the ratio of the interaction strength and the rate of recombination between the loci, is small. Our results not only apply to the transition between genotype and allele selection, but also to localized clusters of interacting genes on the chromosome. Whenever the epistatic fitness difference between different allelic compositions of a cluster exceeds the recombination rate of the cluster, the fittest will amplify exponentially. Because such clusters are often small (36) (one to a few Mb) their recombination rates are low (in the centimorgan range)—hence fitness differentials around 1% can suffice to establish CC dynamics. Selective pressure to reduce recombination load, i.e., the fitness loss through recombination, will therefore favor the evolution of clusters of interacting genes and might be an important driving force for the evolution of

recombination rate (37, 38). The effects described above may provide an explanation for the functional clustering associated with low and high LD regions reported in HapMap (18).

Methods

Random Epistasis Model. A genotype g is described by L binary variables $s_i = \pm 1$, $i = 1, \ldots, L$. To each genotype we assign a fitness

$$F(g) = f \sum_{i}^{L} s_i + \xi(g). \quad [3]$$

The first term is the sum of the additive fitness contributions of the individual loci, each of which has equal magnitude $f = \sqrt{V_A/L}$. The second term is the nonheritable epistatic fitness, where $\xi(g)$ is drawn from a normal distribution with 0 mean and variance V_I. For a uniform distribution of genotypes, the additive fitness variance is V_A, the epistatic variance is V_I, and the total variance is $\sigma^2 = V_A + V_I$.

Pairwise Epistasis Model. Here, we consider epistasis due to pairwise interactions between the different loci. Such pairwise interactions correspond to $s_i s_j$ terms in the fitness function. The fitness of a particular genotype g is determined by the independent effects of the individual loci and the sum of the interactions between all pairs.

$$F(g) = f \sum_{i}^{L} s_i + \sum_{i<j} f_{ij} s_i s_j. \quad [4]$$

When assuming uniform epistasis between all possible pairs, we draw the interaction strength f_{ij} from a Gaussian distribution with 0 mean and variance

$$\frac{2V_I}{L(L-1)}.$$

Clustered Epistasis. To mimic localized clusters of strongly interacting genes on a weakly interacting background, we constructed the matrix of f_{ij}'s as follows. The sparse background epistasis was modeled by assigning each f_{ij} a Gaussian random number with probability $P = 0.1$ and 0 otherwise. Then we built 3 epistatic clusters with centers $c_k = 10, 50, 90$ by adding a Gaussian random number to each f_{ij} with probability

$$p = \exp\left(\frac{(i-c_k)^2 + (j-c_k)^2}{2r^2}\right)$$

with $r = 10$ for $k = 1, 2, 3$. All f_{ij} were rescaled such that $\Sigma_{i<j} f_{ij}^2 = V_I$.

Selection. Our model assumes nonoverlapping generations. In each generation a pool of gametes is produced, to which each individual contributes a number of copies of its genome, which is drawn from a Poisson distribution with parameter $\exp(F(g) - \bar{F})$.

Gene Reassortment. To model gene reassortment in a facultatively mating population, 2 gametes are chosen with probability r and a new genotype is formed by assigning each locus the allele of one or the other parent at random. Otherwise, the new genotype is an exact copy of 1 gamete.

Cross-Overs. Given a cross-over rate ρ per locus, the number of cross-overs is drawn from a Poisson distribution with parameter $(L-1)\rho$ and the cross-over locations are chosen at random. When the number of cross-overs is 0, the offspring inherits the entire genome from 1 parent. To model circular chromosomes, the number of cross-overs is multiplied by 2 enforcing an even number of cross-overs.

Measuring Genetic Diversity. The allele entropy is a convenient descriptor of genetic diversity that is readily calculated from the evolving population. It is defined as $S_A = -\Sigma_i [\nu_i \ln \nu_i + (1-\nu_i)\ln(1-\nu_i)]$, where ν_i is the allele frequency at locus i.

Measuring Linkage Disequilibrium. LD is the deviation of the frequency of a pair of alleles from the random expectation on the basis of the individual allele frequencies, i.e., $D_{ij} = \langle s_i s_j \rangle - \langle s_i \rangle \langle s_j \rangle$. Kimura (21) showed that in QLE

$$\psi_{ij} = \frac{D_{ij}}{\nu_i \bar{\nu}_i \nu_j \bar{\nu}_j}$$

is time independent despite changing allele frequencies ν_i and ν_j ($\bar{\nu}_i = 1 - \nu_i$). To measure genome wide LD, we calculate the sum of all squared LD terms $\Sigma_{i<j} \psi_{ij}^2$. Pairs with ν_i or $\nu_j < 0.01$ or > 0.99 were omitted. A different normalization is used in Fig. 5, where

$$D'_{ij} = \frac{|D_{ij}|}{4 \max(\min(\nu_i \nu_j, \bar{\nu}_i \bar{\nu}_j), \min(\nu_i \bar{\nu}_j, \bar{\nu}_i \nu_j))}$$

is shown (see ref. 19 for a recent review).

ACKNOWLEDGMENTS. We thank Michael Elowitz and Marie-Anne Felix for comments on the manuscript and acknowledge financial support from National Science Foundation Grant PHY05–51164.

1. Begun DJ, et al. (2007) Population genomics: Whole-genome analysis of polymorphism and divergence in Drosophila simulans. *PLoS Biol* 5:2534–2559.
2. Desai MM, Fisher DS (2007) Beneficial mutation selection balance and the effect of linkage on positive selection. *Genetics* 176:1759–1798.
3. Gerrish PJ, Lenski RE (1998) The fate of competing beneficial mutations in an asexual population. *Genetica* 102–103:127–144.
4. McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in Drosophila. *Nature* 351:652–654.
5. Mackay TF (2001) Quantitative trait loci in Drosophila. *Nat Rev Genet* 2:11–20.
6. Barton NH, Keightley PD (2002) Understanding quantitative genetic variation. *Nat Rev Genet* 3:11–21.
7. Brem RB, Storey JD, Whittle J, Kruglyak L (2005) Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature* 436:701–703.
8. Schuldiner M, et al. (2005) Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* 123:507–519.
9. Segrè D, Deluna A, Church GM, Kishony R (2005) Modular epistasis in yeast metabolism. *Nat Genet* 37:77–83.
10. Dobzhansky T (1950) Genetics of natural populations. XIX. Origin of heterosis through natural selection in populations of Drosophila pseudoobscura. *Genetics* 35:288–302.
11. Dolgin ES, Charlesworth B, Baird SE, Cutter AD (2007) Inbreeding and outbreeding depression in Caenorhabditis nematodes. *Evolution* 61:1339–1352.
12. Parker M (1992) Outbreeding depression in a selfing annual. *Evolution* 46:837–841.
13. Edmands S (2008) Recombination in interpopulation hybrids of the copepod Tigriopus californicus: Release of beneficial variation despite hybrid breakdown. *J Hered* 99:316–318.
14. Kuehne HA, Murphy HA, Francis CA, Sniegowski PD (2007) Allopatric divergence, secondary contact, genetic isolation in wild yeast populations. *Curr Biol* 17:407–411.
15. Wright S (1931) Evolution in Mendelian populations. *Genetics* 16:97–159.
16. Hurst LD, Pál C, Lercher MJ (2004) The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet* 5:299–310.
17. Roy PJ, Stuart JM, Lund J, Kim SK (2002) Chromosomal clustering of muscle-expressed genes in Caenorhabditis elegans. *Nature* 418:975–979.
18. International HapMap Consortium (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449:851–861.
19. Slatkin M (2008) Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nat Rev Genet* 9:477–485.
20. Teotónio H, Chelo IM, Bradić M, Rose MR, Long AD (2009) Experimental evolution reveals natural selection on standing genetic variation. *Nat Genet* 41:251–257.
21. Kimura M (1965) Attainment of quasi linkage equilibrium when gene frequencies are changing by natural selection. *Genetics* 52:875–890.
22. Barton NH, Turelli M (1991) Natural and sexual selection on many loci. *Genetics* 127:229–255.
23. Nagylaki T (1993) The evolution of multilocus systems under weak selection. *Genetics* 134:627–647.
24. Franklin I, Lewontin RC (1970) Is the gene the unit of selection?. *Genetics* 65:707–734.
25. Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics* (Longman, Harlow, UK).
26. Lynch M, Walsh B (1998) *Genetics and Analysis of Quantitative Traits* (Sinauer, Sunderland, MA).
27. Hansen TF, Wagner GP (2001) Modeling genetic architecture: A multilinear theory of gene interaction. *Theor Population Biol* 59:61–86.
28. Bulmer MG (1980) *The Mathematical Theory of Quantitative Genetics* (Oxford Univ Press, Oxford).
29. Turelli M, Barton NH (1994) Genetic and statistical analyses of strong selection on polygenic traits: What, me normal? *Genetics* 138:913–941.
30. Eigen M (1971) Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58:465–523.
31. Franz S, Peliti L (1997) Error threshold in simple landscapes. *J Phys A Math Gen* 30:4481–4487.
32. Boerlijst M, Bonhoeffer S, Nowak M (1996) Viral quasi-species and recombination. *Proc R Soc London Ser B* 263:1577–1584.
33. Park J.-M, Deem MW (2007) Phase diagrams of quasispecies theory with recombination and horizontal gene transfer. *Phys Rev Lett* 98:058101.
34. Turelli M, Barton NH (2006) Will population bottlenecks and multilocus epistasis increase additive genetic variance? *Evolution* 60:1763–1776.
35. Rouzine IM, Coffin JM (2005) Evolution of human immunodeficiency virus under selection and weak recombination. *Genetics* 170:7–18.
36. Petkov PM, et al. (2005) Evidence of a large-scale functional organization of mammalian chromosomes. *PLoS Genet* 1:313–322.
37. Barton NH, Otto SP (2005) Evolution of recombination due to random drift. *Genetics* 169:2353–2370.
38. Nei M (1967) Modification of linkage intensity by natural selection. *Genetics* 57:625–641.