Multidimensional Epistasis and the Advantage of Sex

Richard A. Watson
Electronics and Computer Science
University of Southampton
Southampton, SO17 1BJ, U.K
raw@ecs.soton.ac.uk

John Wakeley
Organismic and Evolutionary Biology
Harvard University
Cambridge, MA 02138
wakeley@fas.harvard.edu

Abstract- Kondrashov and Kondrashov (2001) point out that, although common in population genetic models, epistatic systems where the fitness of a genotype is a non-linear function of the number of mutations it carries, which they term "unidimensional epistasis," are a limited subset of possible epistatic landscapes. They claim that more general "multidimensional epistasis" usually confers a disadvantage for sex. However, the evolutionary computation (EC) literature contains models that lie within the space of multidimensional epistasis yet demonstrate an advantage to sex. Here we provide modifications of the Kondrashovs' model that connect with EC results and help to explore the space of epistatic models between the two disciplines.

1 Introduction

Kondrashov and Kondrashov (2001) discuss how simple models of epistasis employed in population genetics are unable to capture potentially important classes of genetic systems. Specifically, the fitness of a genotype is sometimes characterized merely as a function of the number of mutations that the genotype carries (Kondrashov 1982, Gillespie 1994). The number of mutations is a single additive variable – a one dimensional characterisation of all possible genotypes. If each mutation changes fitness by a constant factor then genotype fitness will be an exponential function of the number of mutations, otherwise if the log of fitness is some non-linear function of the number of mutations then the system exhibits epistasis (Shnol & Kondrashov 1993). More generally, a system is epistatic if the fitness effect of a substitution is dependent on genetic background, but by modelling epistasis in a way that is only sensitive to the number of mutations in the genetic background (and not the identity or specific combinations of mutations) it is possible to model epistatic systems without the complication of having mutations at different loci produce different fitness effects (Gillespie 1994).

In general, of course, not all genotypes with the same number of mutations have the same fitness. Even if all mutations have the same fitness effect on average, different combinations of a given number of mutations can have different fitnesses because different loci may have non-uniform epistatic responses to a particular genetic background. Kondrashov and Kondrashov rightly suggest that unidimensional models may be overly restrictive and suggest the term "multidimensional epistasis" to describe epistatic systems where the fitness surface is not a function of any single additive genotype-determined variable.[1]

There are many ways that an epistatic system more general than the single-dimensional kind might be modelled. Kondrashov and Kondrashov first define what they describe as an extreme case where each and every locus is treated as a separate dimension of fitness. Specifically, they describe a mutational path from the ancestral genotype to the maximal fitness genotype that is monotonically increasing in fitness and where there is exactly one beneficial mutation available from each of these genotypes to the next in the path. In this model every locus has a unique set of interactions with the other loci and therefore the exact combination of alleles at all loci must be known in order to determine the fitness of a genotype. The authors also describe a less extreme scenario where the fitness of a genotype is determined by two numbers derived from the number of mutations in two disjoint subsets of loci. A two-dimensional system is a minimal departure from the simple unidimensional model.

Kondrashov and Kondrashov go on to describe simulation results comparing the behaviour of sexual and asexual populations using these examples, and conclude that "unless selection can be approximated by the [unidimensional] fitness potential model, sexual reproduction usually impedes, rather than facilitates, fixation of new, beneficial alleles."

Although unidimensional fitness landscapes, such as one-max (Schaffer & Eshelman 1991), are not uncommon in the theoretical evolutionary computation (EC) literature, general forms of epistasis are the subject of extensive modelling and analysis in EC (e.g. Kauffman 1989, Manderick et al. 1991, Jones 1995, Vose & Wright 1998, Smith & Smith 1999). And the effect of landscapes with complex epistasis on the benefit of sex has been a topic of intense debate (e.g. Holland 1975, Goldberg 1989, Mitchell et al. 1992, Culberson 1995, Stadler & Wagner 1998, Watson et al. 1998, Jansen & Wegener 2001, Spears 1992, 2004, Watson 2001, 2004, 2006). Although the general utility of sexual recombination in evolutionary computation is still debated, many of these works (are designed to) show a benefit to sex and utilise

[1] Note that multidimensional epistasis is not synonymous with non-linear fitness interactions: unidimensional epistasis also requires non-linear fitness interactions.

epistatic systems that are not unidimensional. Moreover, the idea that any algorithmic process or procedural modification such as sex might be universally detrimental or beneficial in general cases is known to be false (Wolpert & Macready 1997). Accordingly, the claim that sex will "usually" confer a disadvantage in systems of multidimensional epistasis would seem unlikely to many in the EC community, and in the next section we modify the Kondrashovs' two-dimensional example to provide some counterexamples to their claim, i.e. cases of multidimensional epistasis that show a strong advantage for a sexual population.

In this light, it would be easy to dismiss the Kondrashovs' result. However, the perspective that they bring to the modelling of epistasis in population genetics should not be underestimated. Some classic population genetic models on the benefit of recombination do not involve epistasis at all (Fisher 1930, Hill and Robertson 1966, Muller 1964); many studies address two-locus two-allele exemplars (see Kondrashov 1993) where the space of possible epistatic structures is severely limited; and the widely-favoured "deterministic mutation hypothesis" (Kondrashov 1982, 1988) is based on a unidimensional model of epistasis. Specifically, the concept of "negative epistasis", a unidimensional epistasis statistic (where log of fitness has a negative second derivative, i.e. two bad mutations together is more than twice as bad as one bad mutation, or two good mutations is less than twice as good as one good mutation) is believed to be a definitive indicator of the benefit of sex (Barton 1995, Feldman et al. 1997, Peters and Otto 2003, Otto et al. 1994). This has motivated several empirical studies hoping to unequivocally settle the question of 'why sex?' (West et al. 1998, Elena and Lenski 1997).

Such statistics of epistasis cannot address the particulars of epistatic interactions among specific combinations of mutations and instead all interactions merely contribute to an average statistic; i.e. on average, is two mutations together more or less fit than expected from the average fitness of one mutation? This view of epistasis as a statistical property of variation in a population lies in contrast to the view, more common in EC, where epistasis is seen as the underlying cause of macro-scale structure in a fitness landscape (Wolf et al. 2000). A statistical treatment, such as negative epistasis, cannot account for a landscape's structural features such as local optima or other restrictions on evolutionary trajectories (Weinreich et al. 2005) that are commonly addressed in EC.

So in this light, a population genetics model that begins to address more sophisticated models of epistasis should be welcomed. Kondrashov and Kondrashov are correct that unidimensional models are not generally predictive of results in more general epistatic landscapes, and that biological epistasis (Phillips et al. 2000, Weinreich et al. 2005), like the epistasis in applied EC problems, requires more sophisticated modelling.

The contrasts of models in population genetics and in EC motivates us to better understand the space of models in between the different types that are popular in the two disciplines. The particular way that Kondrashov and Kondrashov extend into the space of more general epistatic landscapes is not the same as any of the approaches used by the EC examples, and the two-dimensional landscape they describe provides an interesting and simple platform on which to illustrate some of the other (known) effects of recombination in a simple, intuitive and novel manner.

In EC, despite the extensive literature on the subject, and extended theoretical discussion of the different potentialities of recombination, simple and intuitive landscapes that show a fundamental advantage to sex have proved difficult to identify. Recent examples provided by Jansen & Wegener (2001) are simple enough to be amenable to formal proofs of the benefit of sex; but it might be also fair to say that the burden of formal proof, although to its merit, also introduces some complications that obscure intuition in this work. We find that some simple modifications of the Kondrashovs' model can be used to illustrate analogous effects in a straightforward manner. Readers that are already familiar with Jansen & Wegener's work, or similar theory, will not find fundamentally new effects here: we show cases where a sexual population is led to find two different genotypes that cross to make a third fitter genotype that is mutationally distant from either parent. But the models here provide a different, complementary way to illustrate the effects; they show a clear two-dimensional geometric example that is easy to visualise. This helps to build intuition at the same time as building bridges with the population genetics literature.

2 A two-dimensional landscape

The particular example of two-dimensional epistasis that Kondrashov and Kondrashov describe uses epistasis to force a population to follow particular trajectories through genotype sequence space. The paths that are available have only L mutations (where L is the length of the genotype) with no reversions (back mutations) required, but the epistasis places restrictions on the order in which mutations must be accumulated. In the modifications that we investigate in this paper we introduce gaps in the path, requiring multiple point mutations to reach higher fitness points. Kondrashov and Kondrashov suggest that discontinuities in the fitness path will only make the advantage of asexual populations greater, but we find that this is not necessarily the case.

As a specific example of a two-dimensional system the Kondrashovs describe a scenario where there is a narrow path of monotonically increasing fitness that iteratively rewards beneficial mutations first in one subset of mutations and then the other. This system is illustrated in fig. 1a. The fitness of genotype G in this system can be defined using G', where $G'=G$ if G is on the path, and otherwise G' is the closest genotype to G that is on the path. If there is more than one genotype on the path

equally close to G then G' is the highest fitness genotype of this set. Then the fitness of G is defined as $F(G)=1.1^n \times 0.7^d$, where n is the number of 1-alleles in genotype G', and d is the Hamming distance between genotype G and G'. This function rewards genotypes that are on the path with an exponential increase in fitness for each mutation, but penalizes genotypes that stray from the path even though they may have more 1-alleles in some cases. Note that the fitness of a genotype cannot be determined from the total number of mutations it carries but requires a separate count of mutations in the two subsets to determine its proximity to the path.

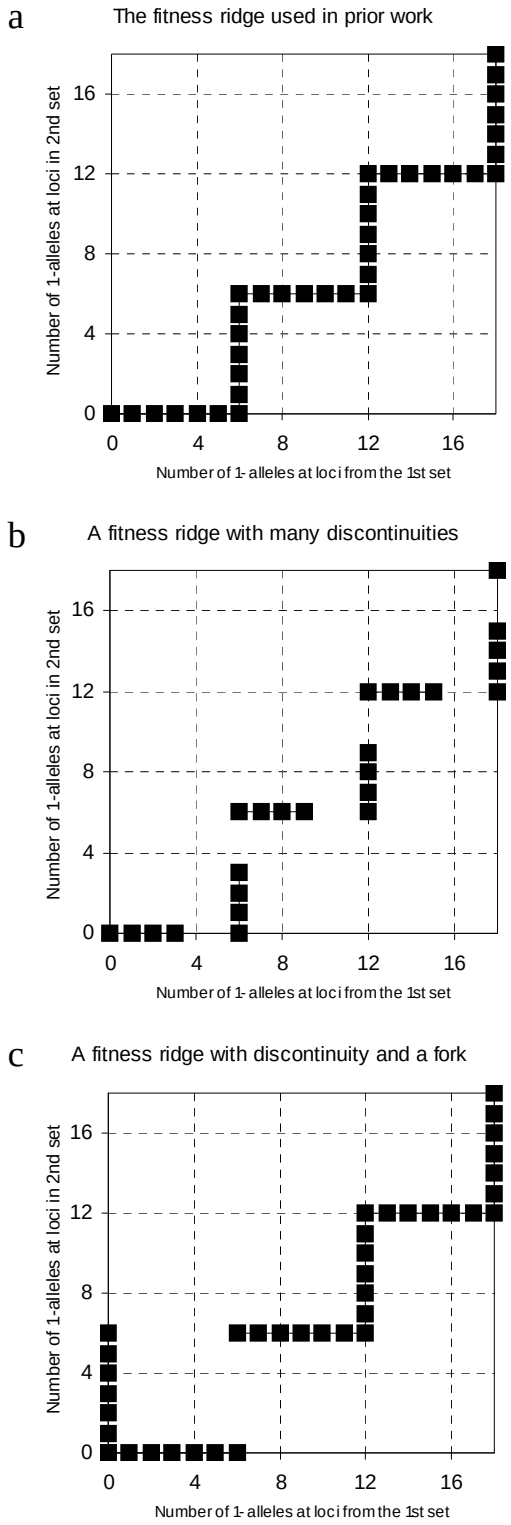

Fig. 1: a) The system described by Kondrashov and Kondrashov as used in their experiments. The fitness ridge connects points (0,0) and (18,18) where a point (p,q) corresponds to any genotype that has p 1-alleles in the first L_1 loci and q 1-alleles at the remaining L_2 loci ($L_1 = L_2 = 18$). b) A fitness ridge with numerous gaps. c) A fitness ridge with a fork.

Kondrashov and Kondrashov use an individual-based simulation to investigate the effect of this epistasis model on the action of recombination. They assume a population of N haploid individuals which undergo unidirectional ($0 \rightarrow 1$) mutation[2] at rate μ per locus. Sex/recombination when used is modelled by random mating followed by free recombination (i.e. *uniform* recombination). Selection is assumed to occur after mutation (and/or recombination).

The authors show that one simulation run of an asexual population fixes all the mutations in the highest fitness genotype (point (18,18) on the path) in approximately half the time used by one run of a sexual population. They observe that the sexual population evolves much faster than an asexual population along straight regions of the continuous fitness ridge but slows down drastically as it approaches a corner. They attribute this delay, at the first corner for example, to the fact that mutations in the second set of loci (i.e. past the corner in the path) are only beneficial when in a genetic background having exactly six 1-alleles in the first set of loci. As the population approaches the corner there are in fact many more than six loci which exhibit 1-alleles in appreciable frequencies, so even if a sexual population sometimes contains fit genotypes that are one mutation around the corner, a mutation in the second set of loci appears deleterious in most backgrounds it encounters at this stage. Thus a sexual population must wait until exactly six loci in the first set approach fixation of the 1-allele and all other loci approach fixation of the 0-allele until a mutation in the second set of loci can rise in frequency in the population. This process occurs slowly given the parameters of the simulation and the exact form of epistasis modelled. In contrast, in an asexual population, should a 1-allele occur in the second set of loci on a particular individual that has six 1-alleles in the first set of loci it will not separated from this background. A resulting genotype can increase in frequency in the population without being disrupted by recombination.

The Kondrashov's conclusion from this result–that multidimensional epistasis usually confers a disadvantage for sex–has influenced theoretical analyses (Ozcelik & Erzan 2003, Soyler & Erzan 2003), and arguments accompanying empirical work (Galvani et al 2003, Kaltz & Bell 2002, West et al 1998). In several cases it is the notion that the order in which mutations occur effects their selective value that has been emphasized. The models the Kondrashovs use do place constraints on the order in

[2] This is a peculiar departure from what might be expected in EC models, but it is retained here to aid comparison.

which mutations may occur (or more exactly, when they will be beneficial) and this is central to the result they describe. This means that although it is clearly implied that the 1-alleles are the 'good' alleles that need to be accumulated to maximize genotype fitness, these alleles are not beneficial in all genetic backgrounds (they thus exhibit sign epistasis, Weinreich et al. 2005). Here we retain the property that the order in which alleles are discovered may be important, but we note that multidimensional epistasis as these authors define it does not require this property. For example, 0→1 mutations are always beneficial in the two-locus two-allele system where the fitness of a genotype, w, is given by $w(0,0)=1$, $w(0,1)=2$, $w(1,0)=3$, $w(1,1)=4$, but w cannot be described as a function of the number of 1-alleles in the genotype.

Below we investigate the robustness of the Kondrashovs' result in the two-dimensional model with respect to changes in population size, mutation rate, and crossover probability. Later we use two relatively small modifications of this model, introducing different types of discontinuities, to illustrate dramatically different effects that may arise in systems of multidimensional epistasis.

3 Investigations on original model

Kondrashov and Kondrashov report that the disadvantage of sex disappears only when the mutation rate becomes lower than $1/N$, forcing non-overlapping allele replacements even in an asexual population. They also state that with $N\mu>1$ differences in alleles frequencies, necessary to initiate selection for a particular set of six 1-alleles at a corner, are small, leading to very slow allele fixations. However, if the disadvantage of sex depends on random fluctuations in allele frequencies being small, then the effect should be sensitive to N (not just $N\mu$): That is, since genetic drift is greater in small populations (Crow & Kimura 1970) the time for a superior combination of mutations to become fixed in a small population may be smaller and thus the delay at corners may be reduced.

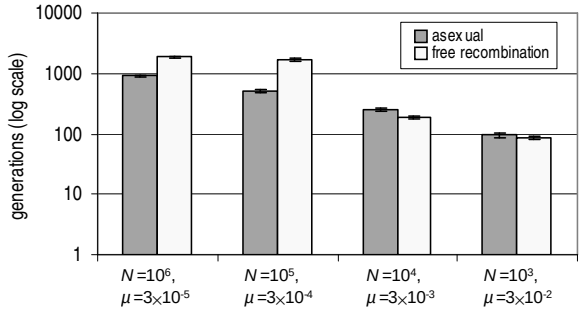

Fig. 2: Sensitivity to N and μ ($N\mu$=30). Showing number of generations for a population to reach fittest genotype – i.e. point (18, 18) occurs with frequency 0.5 or higher.

Fig. 2 shows an exploration of the sensitivity of their simulation to smaller population sizes and higher mutation rates for a constant $N\mu$. Kondrashov and Kondrashov use $N=10^6$, $\mu=3.10^{-5}$. These results utilize a modification of the simulation code provided by Kondrashov and Kondrashov (available on request).[3] Maintaining the same $N\mu$ promotes a significant number of simultaneously segregating mutations in all cases and hence allows the possibility of interference between mutations in one set with mutations in the other. But we see that in some smaller populations with larger mutation rates, specifically $N=10^4$, $\mu=3.10^{-4}$, the result is reversed – i.e. the sexual population has a small advantage, it is approximately 1.3 times faster than the asexual population (a two-tailed Student's t-Test gives a p-value $<< 10^{-5}$). It seems that in this case the benefit of recombination on the straight parts of the path outweigh the disadvantages of sex at the corners of the path–Kondrashov and Kondrashov's observations therefore need to be qualified with respect to values of N and μ. For the remaining experiments we use $N=10^5$, $\mu=3.10^{-4}$ since these parameters show the strongest case of the Kondrashovs' effect of those parameters tested (an asexual population is approximately 3.2 times faster than a sexual population).

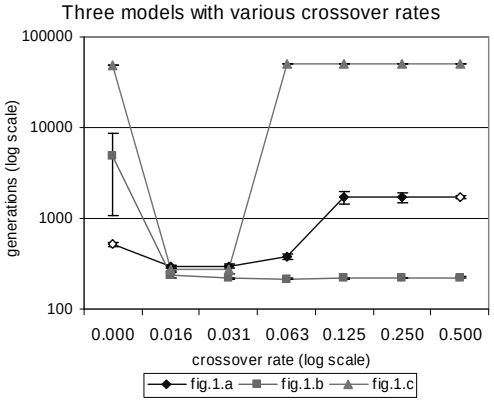

Fig. 3: Mean number of generations for population to reach point (18, 18) on the systems described in fig. 1, showing sensitivity to crossover rate, C. $N=10^5$, $\mu=3\times10^{-4}$. 30 independent simulation runs were conducted for each point. For fig. 1c, settings that fail to reach (18, 18) in 95% or more runs are shown as the max. generation limit = 50,000 generations, other points show the average of runs that succeed. Error bars show +/- one standard deviation. Hollow markers at the ends of the fig. 1a curve indicate points tested by Kondrashov and Kondrashov (here with $N=10^5$). Note that with these simulation parameters, none of the curves are minimized by asexual populations (i.e. $C = 0$).

Fig. 3 provides an exploration of the effect of changing crossover probability, C, for $N=10^5$, $\mu=3.10^{-4}$. C, is the probability of a crossover point occurring between locus i and locus $i+1$; $C=0$ is identical to an asexual population, and $C=0.5$ is free-recombination. Intermediate values of C assume the ordering of loci given by Kondrashov and Kondrashov, i.e. the first set of loci is the first half of the chromosome, and the second set is the remainder. We see that for the system defined by fig. 1a the disadvantage of recombination generally decreases as the amount of

[3] We modified this code to conform to a Wright-Fisher sampling model (Fisher 1930, Wright 1931).

recombination is reduced as expected from the Kondrashovs' results. As we saw in fig. 2 this population size and mutation rate shows a disadvantage of sex (when comparing free recombination with an asexual population). But in fig. 3 we see that there is a combination of parameters where the effect is reversed; a low rate of recombination has a small advantage over an asexual population (1.8 times faster, p-value $<< 10^{-5}$).

Here again it seems that the advantage of sex on the straight sections of the path is greater than the disadvantage of sex at the corners. The important point here is that in two-locus systems or multi-locus systems with uniform epistasis the genetic map and the particulars of the physical linkage in the system are immaterial so the response of a population to changing crossover rate may well be monotonic. But in multi-locus systems with non-uniform epistasis, as here, this cannot be assumed.

4 Discontinuities and free recombination

Kondrashov and Kondrashov assume that the fitness function "forms the narrowest possible continuous ridge connecting the points (0,0) and (18,18)". Actually the path is not very narrow given that there are (*L1* choose *p*)×(*L2* choose *q*) genotypes corresponding to each point (*p*, *q*) on the path – but the path is as narrow as it can be given that it is defined in terms of *p* and *q*. The assumption of a continuous path allows only a very limited class of multidimensional epistatic systems. The authors claim that the disadvantage of sex is even more drastic when a fitness ridge is discontinuous because although a rare double mutant may rise in frequency in an asexual population, in a sexual population its constituent new mutations are deleterious when recombined with existing genotypes.

In their example, Kondrashov and Kondrashov depict a gap at one of the corners in the path, however it can be shown that discontinuities in the straight sections of a path can produce a very different behaviour. Fig. 1b shows a path with many gaps. Unsurprisingly, the progress of an asexual population is significantly delayed by the waiting time for the 3-point mutations required to cross the gaps in this path. But consider free recombination of two different genotypes at the near side of the first gap, i.e. point (3, 0). Recall that there are (18 choose 3)×(18 choose 0)=816 genotypes that map to this point on the path. Many pairs of these genotypes, such as P1 and P2 shown below (spaces indicate the separation of the two subsets of loci), have 1-alleles at disjoint sets of loci :

```
P1: 001000100001000000   000000000000000000
P2: 000000000100100100   000000000000000000
C1: 00?000?00?0??00?00   000000000000000000
```

The result of a recombination between P1 and P2 will produce some offspring genotype C1 which will necessarily have 0-alleles at the loci where both parents had 0-alleles. The probability of the cross producing a genotype at point (6, 0), the closest point on the far side of the gap, is the probability that all the loci where the parents' alleles differ (indicated with "?") receive the 1-allele from the appropriate parent. Under free recombination, each such locus receives the 1-allele with probability 0.5, and thus the overall probability is 2^{-6}. This is higher than the likelihood of crossing this gap by mutation alone which requires changing three loci from 0's to 1's in the first set without changing any other loci in the second set. The probability of this event will depend on the mutation rate but it should be clear that since recombination can, in a sense, focus a high rate of variation on appropriate loci (i.e. those where the parents disagree, Chen 1999) without producing variation in other loci, it provides a clear advantage in this scenario because it increases the likelihood of bringing together these good alleles into one genotype. This ability of a population using free recombination to cross each gap in this path is directly analogous to the advantage of uniform crossover shown in Jansen & Wegener's "gap function" (2001).

This effect requires sufficient diversity in the population such that there are at least two genotypes having exactly three complementary 1-alleles in the first set of loci. Since there is a tendency for populations to converge to a particular set of beneficial loci, this probability may be low in some cases. However, values of N and μ that produce the Kondrashovs' effect are exactly suitable to produce this effect also: their effect arises because competing beneficial alleles take time to resolve and fix a particular block of 1-alleles; our effect arises when different beneficial alleles segregate simultaneously and sometimes produce a useful cross. In both cases, significant population diversity is required to see the effect. Other particulars required for this effect include the fact that in this example the far side of the gap has not more than twice the number of beneficial mutations that the near side of the gap has: this example has the largest gap that can be fitted in a block of 6 loci, but larger systems are less restricted in this respect.

An asexual population is approximately 20 times slower than a sexual population in this particular form of multidimensional epistasis with discontinuities (see fig. 3) despite the fact that this path has just as many 'corners' as fig. 1a. Note also that the time for a sexual population on this system is not only faster than an asexual population but also considerably faster than the time for a sexual population on the ridge without discontinuities. The introduction of gaps in this case reduces the number of competing alleles simultaneously segregating when the population reaches a corner. This shows that discontinuities in the path do not necessarily increase the disadvantage of sex and can produce an advantage.

5 Discontinuities and genetic linkage

The epistatic system in fig. 1b allows diversity to accumulate in the population among genotypes of equal fitness (such as genotypes at point $(0, 3)$) but the production of complementary sets of 1-alleles necessary for a successful cross occurs at random. A much stronger

effect can be produced by a system such as fig. 1c. This system shows a multidimensional epistatic system that differs from the Kondrashovs' system in that it rewards an increase in mutations in the second set in parallel with, rather than after, an increase in mutations in the first set. Thus there are two different ways to increase fitness starting from the ancestral genotype. Crosses between genotypes from different sides of this fork mostly fall into the fitness valley between–a scenario reminiscent of hybrid incompatibility (Orr 1995). However, this model also includes the possibility that some particular (relatively rare) cross of genotypes from the two different branches may be viable: the point (6, 6) in this system is on the path and enjoys the fitness benefit of all twelve mutations. Consider the crossing of a genotype at point (0, 6), with a genotype at point (6, 0). For example:

```
P1: 000000000000000000   010000110000110010
P2: 100110010010001000   000000000000000000
C1: ?00??00?00?000?000   0?0000??0000??00?0
```

The result of a cross between P1 and P2 will produce some offspring genotype C1. The probability of the cross producing a genotype at point (6, 6) is the probability that all remaining loci receive the 1-allele. Under free recombination this is 2^{-12}, which although still better than mutation for the same reasons as before, is very low. However, under a low crossover rate where proximal loci have a higher probability than distal loci of segregating together, the probability is much higher. For example, if there is exactly one crossover point, the probability of a successful cross is at least 1/(L-1) – in the example above it is 5/(L-1) because there are 5 suitable inter-local positions. In contrast, reaching the point (6, 6) from either point (0, 6) or (6, 0) without recombination is highly unlikely: Jumping this gap by mutation requires a 5-point mutation in one set of loci,[4] and at the same time, no change in the number of mutations in the other set of loci. This probability is even lower than the probability of a successful cross with free recombination regardless of the mutation rate used (assuming mutation were bi-directional). This reasoning predicts that both freely recombining populations and asexual populations will do poorly in this system, but sexual populations using a low rate of crossover will cross the gap relatively easily.

Fig. 3 shows the results of simulation runs on fig. 1c for a range of crossover probabilities. The number of successful runs (from 30) for each crossover probability are: C=0.000, 1 run (1300 gens.); C=0.016, 30 runs; C=0.031, 28 runs; C=0.063, 1 run (387 gens.); C=0.125, C=0.250, and C=0.500, 0 runs. For C=0.016, the mean time to point (18, 18) is 271 generations. The runs that fail to reach (18, 18) are tested up to a maximum generation limit of 50,000 generations. Accordingly, in this system of multi-dimensional epistasis, with these parameters, the advantage of sex is more than 184-fold. As predicted this advantage is maximized at a low rate of recombination; in this case C=0.016 (a bit less than 1/L) is superior to both an asexual population and to free recombination. Indeed, a sexual population with a low crossover rate performs as fast as any population does on the fitness path that has no discontinuities. As in the simulations using fig. 1b, surprisingly, it is as though the discontinuity produces no impedance at all to a sexual population with the right amount of crossover, despite being a considerable 'road block' to an asexual population.

This advantage for a sexual population is different from that shown in Section 4 because it exploits the particulars of the genetic map (with low recombination rates) to make larger jumps than those possible with free recombination (note that all recombination rates >0 demonstrate the effect of Section 4, but for the effect in this section, too much sex is no better than none). That is, the effect is sensitive to the ordering of genes on the chromosome–the 1-alleles for each of the two dimensions need to be tightly physically linked so that they travel together during recombination events (Eshelman et al. 1989). If the genetic map is randomised, so that the two-sets of loci are interleaved with one another, the effect is not seen. In Section 4 recombination changes the expected waiting time to cross the gap from being exponential in L to being exponential in the width of the gap, a significant improvement. But here, if the genetic map is favourable, recombination changes the expected waiting time to cross the gap from being exponential in L to being approximately linear in the width of the gap, allowing in principle, much bigger gaps to be crossed and thereby drastically increasing the advantage over asexual populations. Such effects are far more significant than the relatively subtle (approximately two-fold) disadvantages shown by Kondrashov and Kondrashov.

The rigorous treatment provided by Jansen & Wegener (2001) describes the distinction between these advantages of free recombination and low rates of recombination analytically. Our modification of the Kondrashovs' model is, in this respect, just a different example of similar processes. However, a *two*-dimensional epistasis model is ideally suited to illustrating the advantageous distribution of variants that can be produced by crossing *two* parents, and, as such, this minimal departure from unidimensional models helps to build useful intuition for these effects. Moreover, by using a modification of the Kondrashovs' model we bridge between effects described in the population genetic literature and contradictory EC results.

6 Characteristics of natural landscapes

In fig. 1a, although 1-alleles are not beneficial in all backgrounds, there is always at least one beneficial mutation available for every genotype and thus there are no local optima in the fitness surface. Accordingly, finding high-fitness genotypes is not that difficult for any kind of population and the action of sex is at best a subtle one. In contrast, in figs. 1b and 1c a population must escape from the local optima at the edge of the gaps. A

[4] $F(5,6) = F(6,5) = 1.1^{12} \times 0.7 > F(0,6) = F(6,0) = 1.1^{6}$ so a jump that lands one square away from (6, 6) is sufficient for a fitness increase.

population that is able to exploit particulars of the *set* of genotypes that are high-fitness on one side of the fitness gap to find high fitness genotypes on the other side of the gap is much better equipped to make this jump than blind variation from point mutation applied to any one genotype. Accordingly, under appropriate conditions, the advantage of sex is not a subtle one. This effect depends, of course, on how the genotypes on one side of the fitness gap relate to the genotypes on the other. If the genotypes on the higher-fitness side of the gap are arbitrary then recombination will be no better equipped to find them than an appropriate mutation rate. In figs. 1b and 1c the genotypes on the higher-fitness side of the gap are fit because they contain the union of beneficial alleles from two different genotypes on the low-fitness side of the gap–which seems not unreasonable. At the same time, the epistasis in these systems is such that not all combinations of beneficial alleles from different genotypes on the low-fitness side of the gap are superior in fitness–this is what creates the local optima and prevents the efficient action of non-recombining populations.

The Kondrashovs' notion of restrictions on the order in which mutations may occur is certainly biologically plausible, but the exact model that they use to investigate the effect of this assumption is only one of many possibilities. If we are willing to suppose a narrow fitness path where the number of beneficial mutations that are available at any one time is limited, it is a small step to a scenario where for some genotypes there are *no* beneficial (single-point) mutations available, creating local optima. Kondrashov and Kondrashov are open to the possibility of discontinuities in the path in their own discussion, but here, in the first case at least, we simply position these discontinuities differently. Our second case (fig. 1c) is built from the notion of crosses between diverging populations. It is quite plausible to suppose that diverging subsets of a population, especially in a subdivided population, might arrive at different, largely incompatible, regions of genotype space. Here however, we are adding the assumption that some rare cross between these types is fit. Like the Kondrashovs' models, our models are only one possible way to model such assumptions. They are nonetheless sufficient to provide counterexamples for the Kondrashovs' general claim.

We have shown several counterexamples where sex provides an advantage in systems of multidimensional epistasis–even using the Kondrashovs' own specific form of multidimensional epistasis shown in fig. 1a. Their claim that multidimensional epistasis usually confers a disadvantage for sex is therefore not well supported. Similarly their claim that "unless data will show that ridges of high fitness are mostly straight and rarely contain corners, facilitation of adaptive evolution cannot be the reason for the origin and maintenance of sex" appears to be incorrect. However, despite having examples that counter the Kondrashovs' results, we certainly do not claim that sex usually provides an advantage in systems of multidimensional epistasis. The actual properties of natural epistatic systems and the benefit or disadvantage of sex in such systems is an empirical matter yet to be determined. Meanwhile, the systems described in these simulations serve to show that multidimensional epistasis can have varied effects: showing advantages as well as the disadvantage shown by Kondrashov and Kondrashov.

7 Conclusions

The Kondrashovs' model of two-dimensional epistasis has proved very useful in reinforcing an important general point–that the unidimensional model represents a highly restricted class of systems, and results pertaining to the benefit or disadvantage of sex under the unidimensional epistatic model are not predictive of the benefit or disadvantage of sex in general epistatic systems. Classifications such as negative and positive epistasis are not as comprehensive as they might first appear: whether the fitness of genotypes tends on average to increase faster or slower than the exponential of the number of mutations they carry does not reveal epistatic effects that restrict evolutionary trajectories, for example. Accordingly the impact of empirical investigations assessing epistasis in such simplistic frameworks should not be overstated.

The two-dimensional system of Kondrashov and Kondrashov provides a valuable means to illustrate two effects that are not accommodated by simplistic models of epistasis. Both cases show the ability of a population to jump across a fitness gap: one case (using fig. 1b) utilizes free recombination; the other (using fig. 1c) uses a low per locus rate of crossover. In both cases the action of recombination facilitates a jump from a local fitness peak to a genetically distant point of higher fitness–i.e. several point mutations are required to find a point of higher fitness. Certainly, a population without recombination can make such jumps with non-zero probability and this probability may be optimized by careful tuning of the mutation rate. But from the reasoning provided above and elsewhere (Jansen & Wegener 2001) we can see that the probability of a recombining population making this jump, given an appropriate pair of parents and a low mutation rate, is higher than the highest probability of making this jump from either parent by spontaneous point mutations alone under any mutation rate. The simulations confirm that there is some combination of parameters where the population can provide appropriate pairs of parents in these systems.

The divide between the types of models considered reasonable in population genetics and in evolutionary computation is a large one. In this paper we have used modifications of the Kondrashov's model to illustrate some effects known in EC. These provide new intuitive illustrations of the effects and utilise the groundwork laid for a population genetic audience. Future work must continue to populate the space of models in this divide.

Acknowledgments. We are especially grateful to Fyodor Kondrashov for providing simulation code, and to the anonymous reviewers for their detailed comments.

Bibliography

Barton N.H., 1995, "A general model for the evolution of recombination", *Genet. Res*. **65**: 123–144.

Chen, S., 1999, *Is the Common Good? A New Perspective Developed in Genetic Algorithms*, PhD diss., Robotics Institute, Carnegie Mellon University.

Crow, J.F. and M. Kimura, 1970, *An introduction to population genetics theory*, Harper and Row, New York.

Culberson, J.C., 1995, "Mutation-Crossover Isomorphisms and the Construction of Discriminating Functions", *Evolutionary Computation*, **2**, 279–311.

Elena, S.F., & R.E. Lenski, 1997, "Test of synergistic interactions among deleterious mutations in bacteria", *Nature* **390**: 395–398.

Eshelman, L.J., R.A. Caruana, & J.D. Schaffer, 1989, "Biases in the crossover landscape". *Proc. 3rd Int'l Conference on Genetic Algorithms*, J.D. Schaffer (ed), 10–19.

Feldman, M.W., S.P. Otto & F.B. Christiansen, 1997, "Population genetic perspectives on the evolution of recombination" *Annu. Rev. Genet.* **30**: 261–295

Fisher, R.A., 1930. *The genetical theory of natural selection*. Dover, New York.

Galvani, A.P., Coleman, R.M., & Ferguson, N.M., 2003, "The maintenance of sex in parasites", *P. Roy. Soc. Lond. B. Bio.* **270** (1510), 19–28.

Gillespie, J.H., 1994, "Substitution processes in molecular evolution. II. Exchangeable models from population genetics" *Evolution* **48**:1101–1113.

Goldberg, D.E., 1989, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley.

Hill, W.G., & A. Robertson, 1966, "The effects of linkage on limits to artificial selection", *Genet. Res.* **8**:269–294

Holland, J.H., 1975, *Adaptation in Natural and Artificial Systems*, Ann Arbor, MI: The University of Michigan Press.

Jansen, T., & L. Wegener, 2001, "Real Royal Road Functions - Where Crossover Provably is Essential", in Spector et al. (eds), *Procs. of Genetic and Evolutionary Computation Conference*, Morgan Kaufmann, San Francisco, CA. 374–382.

Jones, T, 1995, *Evolutionary Algorithms, Fitness Landscapes and Search*, PhD dissertation, 95-05-048, University of New Mexico, Albuquerque.

Kaltz, O., & G. Bell, 2002, "The ecology and genetics of fitness in Chlamydomonas. XII. Repeated sexual episodes increase rates of adaptation to novel environments", *Evolution* **56** (9), 1743–1753.

Kauffman, S.A., 1989, "Adaptation on Rugged Fitness Landscapes", *Lectures in the Sciences of Complexity*, Vol 11, 527–618. Addison Wesley.

Kondrashov, A.S., 1982, "Selection against harmful mutations in large sexual and asexual populations", *Genet. Res.* **40**: 325–332.

Kondrashov, A.S., 1988, "Deleterious mutations and the evolution of sexual reproduction", *Nature* **336**:435–440.

Kondrashov, A.S., 1993, "Classification of hypotheses on the advantage of amphimixis", J. Hered. 84,372-387

Kondrashov, F.A. & Kondrashov, A.S., 2001, "Multidimensional epistasis and the disadvantage of sex", *PNAS*, **98**:21, 12089–12092.

Manderick, B., M. de Weger, & P. Spiessens, 1991, "The genetic algorithm and the structure of the fitness landscape", *4th International Conference on Genetic Algorithms*, R.K. Belew, L.B. Booker (eds.), Morgan Kaufmann, 143–150.

Mitchell, M, S. Forrest & J.H. Holland, 1992, "The royal road for genetic algorithms: Fitness landscapes and GA performance", *Procs. First European Conference on Artificial Life*. MIT Press. 245–54.

Muhlenbein, H., 1992, "How genetic algorithms really work: I. mutation and hill-climbing." *Parallel Problem Solving from Nature 2*, R. Manner & B. Manderick (eds), Elsevier. 15–25.

Muller, H.J., 1964, "The relation of recombination to mutational advance", *Mutat. Res.* **1**:2–9

Orr, H.A., 1995, "The Population Genetics of Speciation: The Evolution of Hybrid Incompatibilities", *Genetics* **139**: 1805–1813.

Otto, S.P., M.W. Feldman, & F.B. Christiansen, 1994, "Some advantages and disadvantages of recombination", *Lecture Notes in Biomathematics*, S.A. Levin, (ed). 198–211. Springer-Verlag.

Ozcelik, S., & Erzan, A., 2003, "Hamiltonian model for multi-dimensional epistasis, *Int. J. Mod. Phys. C.* **14** (2), 169–173.

Peters, A.D. & S.P. Otto, 2003, "Liberating genetic variance through sex" *BioEssays*, **25**: 533–537.

Phillips, P.C., S.P. Otto, & M.C. Whitlock, 2000, "Beyond the average: The evolutionary importance of epistasis and the variability of epistatic effects", in Wolf et al. 2000.

Schaffer J.D. & L.J. Eshelman, 1991, "On crossover as an evolutionary viable strategy", in R.K. Belew & L.B. Booker, editors. *Proceedings of the 4th International Conference on Genetic Algorithms*, Morgan Kaufmann. 61–68.

Shnol, E.E. & Kondrashov, A.S., 1993, "The Effect of Selection on the Phenotypic Variance", *Genetics* **193**: 995–996.

Smith, R.E., & J. Smith, 1999, "An Examination of Tunable, Random Search Landscapes", *Proceedings of the Fifth Workshop on Foundations of Genetic Algorithms*, W. Banzhaf, C.R. Reeves (eds.) Morgan Kaufmann. 165–181

Soyler, G., & Erzan, A., 2003, "Equilibrium correlations in a model for multidimensional epistasis", *Theor. Biosci.* **121** (4), 387–394.

Spears, W.M., 1992, "Crossover or Mutation?", *Foundations of Genetic Algorithms-2*, D. Whitley (ed.). 221–237.

Spears, W.M., 2004, *Evolutionary Algorithms: The Role of Mutation and Recombination*, Springer-Verlag.

Stadler P.F., & G.P. Wagner, 1998, "Algebraic theory of recombination spaces", *Evolutionary Computation*, **5**, 241–275.

Vose M.D., & A.H. Wright, 1998, "The simple genetic algorithm and the Walsh transform: Part I, Theory, *Evolutionary Computation*. **6**(3):253–73.

Watson, R.A., 2004, "A Simple Two-Module Problem to Exemplify Building-Block Assembly Under Crossover", in *Parallel Problem Solving from Nature (PPSN VIII)*, Xin Yao, et al. (eds.), LNCS 3242, Springer, 161–171.

Watson, R.A., 2006, *Compositional Evolution: The impact of Sex, Symbiosis and Modularity on the Gradualist Framework of Evolution*, MIT Press, to appear.

Watson, R.A., G.S. Hornby & J.B. Pollack, 1998, "Modeling Building block Interdependency", *Parallel Problem Solving from Nature V*, A. Eiben, et al. (eds.), Springer. 97–106.

Weinreich, D.M., R.A. Watson, & L. Chao. 2005. Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **59** (6):1165–1174.

West S.A., Peters A.D., & Barton N.H., 1998, "Testing for epistasis between deleterious mutations", *Genetics* **149** (1): 435–444.

Wolf, J.B., E.D. Brodie, & M.J. Wade, 2000, *Epistasis and the evolutionary process.* Oxford Uni. Press.

Wolpert, D, & W. Macready, 1997, "No Free Lunch Theorems for Optimization", *IEEE Transactions on Evolutionary Computation*, **1** (1):67–82.