PROCEEDINGS OF THE ROYAL SOCIETY B
FirstCite® e-publishing

Proc. R. Soc. B
doi:10.1098/rspb.2005.3338
Published online

Sexual reproduction reshapes the genetic architecture of digital organisms

Dusan Misevic[1,*], Charles Ofria[2] and Richard E. Lenski[1,3]

[1]*Ecology, Evolutionary Biology and Behavior Program,* [2]*Department of Computer Science and Engineering, and* [3]*Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824, USA*

Modularity and epistasis, as well as other aspects of genetic architecture, have emerged as central themes in evolutionary biology. Theory suggests that modularity promotes evolvability, and that aggravating (synergistic) epistasis among deleterious mutations facilitates the evolution of sex. Here, by contrast, we investigate the evolution of different genetic architectures using digital organisms, which are computer programs that self-replicate, mutate, compete and evolve. Specifically, we investigate how genetic architecture is shaped by reproductive mode. We allowed 200 populations of digital organisms to evolve for over 10 000 generations while reproducing either asexually or sexually. For 10 randomly chosen organisms from each population, we constructed and analysed all possible single mutants as well as one million mutants at each mutational distance from 2 to 10. The genomes of sexual organisms were more modular than asexual ones; sites encoding different functional traits had less overlap and sites encoding a particular trait were more tightly clustered. Net directional epistasis was alleviating (antagonistic) in both groups, although the overall strength of this epistasis was weaker in sexual than in asexual organisms. Our results show that sexual reproduction profoundly influences the evolution of the genetic architecture.

Keywords: AVIDA; epistasis; experimental evolution; modularity; recombination

1. INTRODUCTION

Modules are ubiquitous in biological systems, and they appear to be a critical aspect of biological organization (Hartwell *et al.* 1999; Schlosser & Wagner 2004). Defined as groups of characters serving the same function that are integrated into a unit largely independent from other such units (Wagner 1996), modules occur in such diverse contexts as the *Hox* gene cluster (Ferrier & Holland 2001), butterfly wing development (Beldade *et al.* 2002) and metabolic networks in bacteria and eukaryotes (Ravasz *et al.* 2002). Modular organization of genomes may facilitate the exchange of independent 'building blocks' via recombination, increase phenotypic variability and promote evolvability (Kirschner & Gerhart 1998). While sexual reproduction with obligatory recombination is also common (Bell 1982), the selective forces responsible for its origin and maintenance—in the face of substantial costs—remain largely unknown even after a century of discussion and investigation (Williams 1975; Maynard Smith 1978; Michod & Levin 1988; Kondrashov 1993; West *et al.* 1999). One hypothesis emphasizes genetic architecture and suggests that aggravating (synergistic) epistasis between deleterious mutations may favour sex (Kondrashov 1982; Wolf *et al.* 2002), but various experiments show no excess of aggravating relative to alleviating epistasis (Chao 1988; de Visser *et al.* 1997; Elena & Lenski 1997; Lenski *et al.* 1999; Wilke *et al.* 2003). While these and other studies have considered the effects of genetic architecture on the evolution of sex (Kondrashov 1982; Rice & Chippindale 2001), the effects of reproductive mode on genetic architecture have received less attention (but see Malmberg 1977; Lawrence & Roth 1996; Lenski *et al.* 1999; Pál & Hurst 2003, 2004). Here, by contrast, we measure both modularity and epistasis in the genomes of sexual and asexual evolving computer programs. We expect sex to promote more modular genomes, which may accelerate the origin of new traits and avoid disrupting existing traits. We expect sexual and asexual reproduction to yield different patterns of epistasis for two reasons. First, sex favours those mutations that enhance fitness across different genetic backgrounds, while asexuality favours mutations that are beneficial in the background in which they occur. Thus, stronger epistatic tendencies may evolve in asexual rather than in sexual reproduction (Malmberg 1977). Second, sex may promote aggravating epistasis relative to alleviating epistasis, because recombination would then facilitate the efficient removal of deleterious mutations. This explanation has been proposed for the evolution of sex (Kondrashov 1982), but the causal link might also be reversed.

In this study, we used 'digital organisms' to examine the effects of reproductive mode on the evolution of genetic architecture. These digital organisms are computer programs that replicate, mutate and evolve in populations maintained by the AVIDA software (Wilke & Adami 2002; Ofria & Wilke 2004). They can perform various functions by executing the series of instructions encoded in their genomes, including instructions that enable them to copy their genomes line by line and thereby reproduce. Depending on the genetic program encoded by their genome, instructions may be executed out of order or multiple times. Point mutations, insertions and deletions occur randomly during this process. Organisms compete for the energy they need to execute their genomic

* Author for correspondence (dule@alife.org).

Received 20 June 2005
Accepted 20 September 2005

programs, and the resulting selection acts on heritable differences in their performance that are generated by mutations and, in sexual populations, by recombination. Evolution therefore modifies the genome, with selection tending to reduce the number of instructions that must be executed to reproduce while increasing the energy available for execution. Organisms can augment the basal energy obtained at birth by performing computations in a manner analogous to metabolizing resources. In this study, nine distinct resources occur in unlimited quantities in the environment, but an individual can make use of each resource only once during its life cycle. The ancestor cannot perform any of these computations, but populations can evolve the ability to perform them.

Several studies of evolutionary dynamics and outcomes have taken advantage of the speed of evolution, flexible experimental design and extensive data that can be obtained with Avida (Lenski *et al.* 1999, 2003; Adami *et al.* 2000; Wilke *et al.* 2001; Chow *et al.* 2004; Misevic *et al.* 2004). Here, we examine the evolution of two key features of genetic architecture—modularity and epistasis—as a function of reproductive mode. To do so, we extended Avida to allow the possibility of sexual reproduction. Following the requisite site-by-site copying of a genome, the asexual divide instruction performs a genomic division and places the offspring into the population. The new sexual divide instruction requires the exchange of genetic material between two separately copied genomes before the recombinant offspring are placed in the population. Therefore, all offspring are the product of recombination under the sexual regime in Avida. We used two ancestors, capable of self-replication but not of performing any computations, and differing only in their divide instruction, to seed 100 sexual and 100 asexual populations in identical environments. Such experiments and subsequent analyses simply cannot be performed with any organic system at the scale, scope and precision that digital organisms allow. We realize, of course, that the detailed results of similar experiments and analyses would undoubtedly differ between digital and organic systems (as they probably would also for different organic systems, if such work could be performed). Our intent, however, is to test general hypotheses about genome architecture in relation to mode of reproduction.

2. MATERIAL AND METHODS

(a) *The Avida system*

We performed experiments with the Avida software, which is available without cost at http://devolab.cse.msu.edu/software/avida/. All experiments used default settings unless otherwise indicated. Reproductive mode is determined by whether the divide-sex or divide-asex instruction is included in the instruction set. All genomes used the default set of 25 instructions plus one of these divide instructions. Point, insertion and deletion mutations occurred at rates of 0.002, 0.0005 and 0.0005 per instruction copied, respectively. Point mutations swap an instruction in the genome with a random one from the instruction set, while insertion and deletion mutations add or delete a random instruction. Besides their mode of reproduction, populations differed only in a random number seed that affected all stochastic events during an experiment, including mutations and offspring placement. Each population had a maximum size of 3600 organisms and was propagated for 100 000 updates. An update equals the time in which 30 instructions, on average, are executed per organism in the population. A typical generation is 5–10 updates, with the exact value depending on the complexity of organismal phenotypes, which changed during evolution. With asexual reproduction, an offspring is created by a division and immediately placed in the population. In sexual reproduction, the genome of an incipient offspring is first placed in a 'birth chamber' following a division. If the chamber is empty, the genome remains there until a second one arrives. (Sexual reproduction thus introduces a delay associated with the requirement for pairing. In principle, this delay might slow evolution in sexual populations relative to asexual ones. Pilot experiments in which some asexual populations experienced a corresponding delay, while others did not, showed that this delay had no appreciable effect on the rate of adaptation.) When two genomes are present, they recombine by swapping a continuous region of code. The relative locations of the beginning and end of that region are chosen from a uniform random distribution in the 0–1 range; the absolute locations are calculated by multiplying the random numbers by genome lengths. Genomes are circular and sites are numbered starting with the first one executed after birth. After recombination, both offspring are placed at random locations in the population. Organisms have no means of mate choice, and thus sexual selection is absent from this system. For both reproductive modes, offspring placement kills the organism that previously occupied that location, introducing genetic drift while maintaining a constant population size. All new organisms receive a basal amount of energy, which is scaled by genome length, thereby eliminating direct selection for small genomes. If, in the course of executing its genomic program, an organism inputs one or two 32 bit strings and outputs the result of one of nine basic logic operations performed in a bit-wise fashion on those strings, then it obtains additional energy at a rate roughly proportional to the operation's complexity. There are many different ways to perform each operation, and the number and identity of instructions used vary among organisms. The energy obtained by this computational metabolism is combined multiplicatively with the basal energy, and the product determines the relative speed with which each organism executes its genomic program. An organism's expected fitness is calculated as its total energy divided by the time used to produce an offspring and corresponds to the rate of offspring production. Fitness is expressed relative to the common ancestor; in our analyses of evolved populations, relative fitness was averaged over all organisms from a population and then log transformed.

(b) *Modularity*

We obtained an organism's genotype–phenotype (GP) map by separately mutating each site in its genome and recording any resulting changes in the computational traits performed by the mutants. Physical deletions often produce confounding effects caused by changes in genome length; therefore, we mutated sites to an inert instruction (outside the default set available during evolution) that acts as a placeholder only. The GP map identifies those genomic regions that encode different traits. From the GP map, we then calculated two indices, physical (PM) and functional modularity (FM).

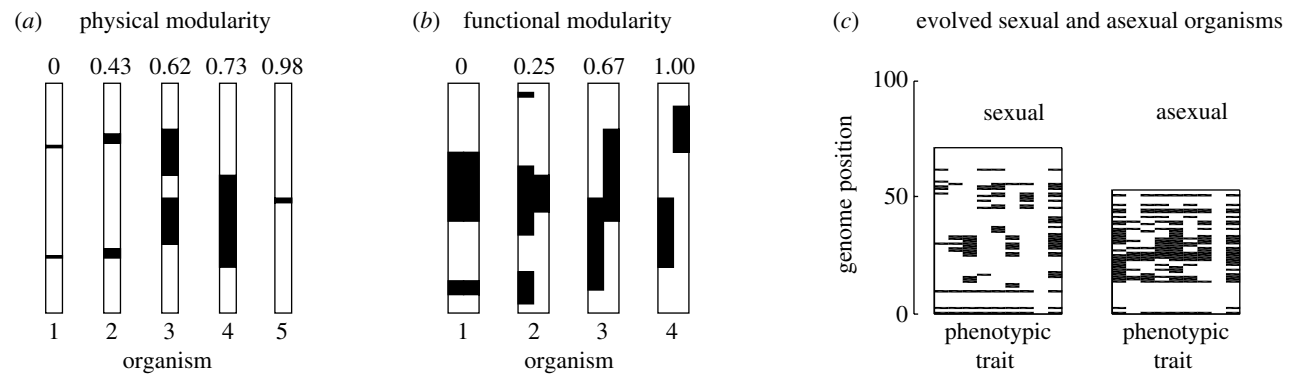

Figure 1. Genotype–phenotype maps illustrating physical and functional modularity. Phenotypic traits are arrayed as columns and genomic sites as rows. For simplicity, the circular genomes are shown in a linear fashion. An open cell indicates that the instruction at that site can be deleted without eliminating the trait; a filled cell marks the site of an instruction required for expression of the trait. If a mutation not only disrupts a trait but also prevents the organism from replicating, then that site is not filled in the GP map. (*a*) Five hypothetical organisms of equal length that express a single functional trait, with physical modularity values indicated above. (*b*) Four hypothetical organisms of equal length that express two traits each, with functional modularity values indicated above. (c) Representative GP maps for sexual (left) and asexual (right) organisms. PM and FM values for this sexual organism are 0.796 and 0.841, respectively; PM and FM values for the asexual organism are 0.689 and 0.746, respectively. Each value is within 5% of the overall mean for the corresponding reproductive mode.

PM measures the mean distance between sites encoding all computational traits

$$\mathrm{PM} = 1 - 2\frac{\sum_{t\in T}\sum_{i,j\in S_t, i\neq j}\frac{d(i,j)}{n(S_t)(n(S_t)-1)}}{L\times n(T)},$$

where L is genome length, T is the set of an organism's traits, $n(T)$ is the number of traits, S_t is the set of all sites encoding trait t, $n(S_t)$ is the number of sites encoding trait t and $d(i,j)$ is the distance between sites i and j. The index averages the distance between two sites encoding a trait over the number of site pairs for each trait and the number of traits, and normalizes by genome length, which allows comparisons between organisms that differ in length and number of expressed traits. The normalized average is then doubled and subtracted from unity, yielding a score from 0 to 1, with higher values indicating more modular architectures. Characteristic distributions of instructions within a trait and the corresponding PM values are illustrated using hypothetical organisms in figure 1*a*; evolved digital organisms generally require many more instructions to perform functional traits than these examples, which serve only to illustrate features of the PM metric. Organism 1 has the lowest possible modularity because the trait is encoded by only two instructions that are located as far apart from each other as possible given a circular genome. Organism 2 is more modular than 1 because, while the relevant instructions also lie in two distant regions, the multiple instructions within each region are very close together and, thus, lower the average distance. Organism 4 is more modular than 3 because all of the relevant instructions in 4 are within a single genomic region. Organism 5 has the highest PM because the trait is encoded by just two adjacent instructions. While the PM for Organism 5 is not quite equal to the theoretical upper bound of 1, it would asymptotically approach that bound with increasing genome length.

FM measures the average overlap in the genomic sites that encode two or more different traits

$$\mathrm{FM} = \frac{\sum_{a,b\in T, a\neq b}\sum_{s\in S_a}(1-m(s,b))}{Ln(T)(n(T)-1)},$$

where $m(s,b)$ specifies whether site s is required for expression of trait b, with $m(s,b) = \{1 \text{ if } s\in S_b;\ 0 \text{ if } s\notin S_b\}$. The number of non-overlapping sites for two traits is averaged over all trait pairs and normalized by genome length, again allowing comparisons between organisms that differ in length and number of expressed traits. Characteristic types of overlap between traits and the corresponding FM values are shown for hypothetical organisms in figure 1*b*. When there is complete overlap between the instructions encoding different traits, then FM is 0, as in Organism 1. By contrast, FM equals 1 when the instructions encoding different traits have no overlap at all, as in Organism 4. Intermediate values of FM are represented by Organisms 2 and 3. Organisms that do not express any computational traits (including those sampled before any functions evolved and, later, those with severe mutations) have undefined PM and were excluded from the analysis. Similarly, FM is defined only for organisms expressing at least two functions because it measures the overlap in the instructions that encode them.

Modularity has joined the ranks of biological properties, such as fitness and species, that are often difficult to define and measure, resulting in a multitude of different and sometimes even conflicting definitions (Wagner 1996; Hartwell *et al.* 1999; Ferrier & Holland 2001; Winther 2001; Beldade *et al.* 2002; Ravasz *et al.* 2002; Schlosser & Wagner 2004). In pilot experiments, we explored several different modularity metrics ranging from quite simple (e.g. average number of tasks affected by an instruction) to more complex ones (e.g. FM further normalized by the expected overlap given the density of instructions in the GP map). Results obtained with all the metrics led to qualitatively similar conclusions. Therefore, we chose to focus on PM and FM, which measure different, yet intuitive and meaningful aspects of genomic architecture.

(c) *Mutational sensitivity and epistasis*

We mutated each site in an organism's genome to every alternative state (e.g. $50\times(26-1)=1250$ mutants for length 50 genomes and 26 different possible instructions) and measured the resulting fitness. For each organism, we also examined random samples of a million mutants for each mutational distance from 2 to 10; these samples included

Table 1. Comparisons of properties between sexual and asexual evolved populations.
(The p-values are based on two-tailed t-tests.)

response variable	mean sexual (±s.d.)	mean asexual (±s.d.)	p
log (fitness)	5.762 (1.610)	5.198 (1.228)	0.006
log (genome length)	2.022 (0.414)	1.853 (0.114)	<0.001
physical modularity, PM	0.824 (0.098)	0.699 (0.073)	<0.001
functional modularity, FM	0.838 (0.091)	0.766 (0.079)	<0.001
average effect of single point mutations on fitness, α	0.546 (0.292)	0.768 (0.190)	<0.001
net directional epistasis, β	0.929 (0.057)	0.862 (0.088)	<0.001

more than 18 billion mutants in all. Parameters α and β define an organism's average sensitivity to individual mutations and the overall form of epistatic interactions among mutations, respectively (Elena & Lenski 1997; Lenski *et al.* 1999). We calculated each organism's α exactly as $-\log W(1)$, where $W(1)$ is the average fitness of all single mutants expressed relative to the unmutated state. We then estimated each organism's β by minimizing the sum of squared deviations between the actual average fitness values for $1 \leq m \leq 10$ and those predicted by the power function $\log W(m) = -\alpha m^{\beta}$, using α as obtained above.

(d) *Statistics*

We used SYSTAT v. 10.2 software (SSI, Richmond, CA) for all statistical tests.

3. RESULTS

(a) *Evolved sexual organisms have higher fitness and longer genomes*

Sexual populations adapted significantly better to the environment, on average, than did asexual populations, as indicated by average fitness values after 100 000 updates ($p = 0.006$; table 1; figure 2*a*). Sexual populations also evolved larger genomes, on average, than asexual populations ($p < 0.001$; table 1; figure 2*b*), but this difference was strongly influenced by the fact that 35 sexual populations evolved genomes more than twice the ancestral length of 50, whereas only nine asexual populations did so. Similarly, there was more diversity in genome length within individual sexual populations (data not shown), probably caused by the larger changes in genomes allowed by recombination. The large genomes typically evolved via genome doublings that occurred when, owing to some mutation, the genetic program failed to detect that the genome had already been copied and repeated the site-by-site replication. Sexual reproduction in AVIDA evidently increased the chance of genome doubling, created more favourable genetic combinations in the larger genomes, or both. To preclude genome doublings from possibly biasing our analyses of genetic architecture, we reduced the dataset by excluding all populations with mean length greater than or equal to 100 (i.e. at least twice the ancestral length; see electronic supplementary material). We also performed additional experiments in which we actively prevented genome doublings from occurring (see electronic supplementary material). In both cases, the resulting asexual populations had, on average, longer genomes than sexual ones, thus reversing the direction of potential bias. The alternative analyses also eliminated the significant fitness difference between sexual and asexual populations, with higher fitness shifting, albeit insignificantly, to the asexual populations. Importantly, all of the qualitative effects of reproductive mode on genetic architecture remained in the same direction and significant in all but one case, under both alternative analyses. Therefore, these architectural differences are robust with respect to differences in genome length between sexual and asexual organisms.

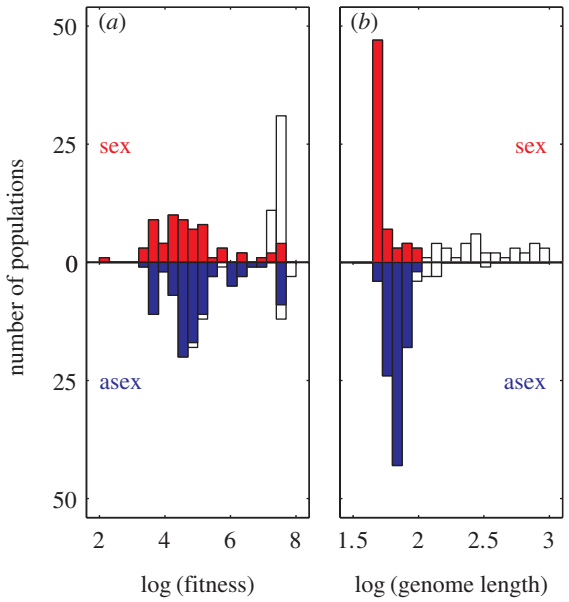

Figure 2. (*a*) Final distributions of average fitness and (*b*) genome length in sexual and asexual populations, respectively. Open sections show populations with average genome length greater than or equal to 100 instructions.

(b) *Evolved sexual organisms have more modular genomes*

We performed all of the analyses of genetic architecture on a random sample of 10 viable organisms from each population at the end of the evolution run; occasional non-viable genotypes were excluded. For each organism, we first deleted each genomic site (replacing it with an inert placeholder instruction) in order to construct GP maps; these maps identify the set of sites required to express a particular trait (compute a logic operation). We used the independently evolved populations as the unit of replication in statistical comparisons, because organisms sampled from the same population inevitably share much of their ancestry.

Using the GP maps, we then calculated PM, which reflects the average distance between two sites encoding a particular trait, and FM, which captures the average overlap between sites encoding different traits. PM and FM can range from 0 to 1, with high PM values indicating that traits are encoded in compact regions of the genome and high FM values corresponding to low overlap between traits. Representative GP maps illustrate the difference in modularity between sexual and asexual organisms: sites encoding the various traits in sexual organisms tend to be more compact (shorter, more continuous vertical filled blocks) and less overlapping (shorter, fewer horizontal filled lines) than those in asexual organisms (figure 1). We averaged PM and FM across organisms within a population, and then compared the populations with different reproductive modes. Over time, the genomes of sexual organisms became increasingly more physically modular than asexual ones (figure 3*a*), and this difference was highly significant at the end of the experiment ($p<0.001$; table 1; figure 3*b*). The difference in PM remained highly significant when populations with genome length greater than or equal to 100 were excluded from the analysis ($p<0.001$; see electronic supplementary material). Similarly, sexual organisms evolved genomes with a significantly higher FM than asexual ones ($p<0.001$; table 1, figure 3*c*), a difference that also holds in the reduced dataset ($p=0.002$; see electronic supplementary material).

(c) *Mutational sensitivity and epistasis differ between sexual and asexual organisms*

We further examined genetic architecture by quantifying the form and extent of epistasis in the same organisms used to assess modularity. For each organism, we made all possible one-step point mutants and obtained random samples of a million organisms carrying from 2 to 10 mutations; for each mutant, we measured its fitness relative to its unmutated parent. We averaged relative fitness over the organisms in a mutational class from each population. We then used a power function, $\log W(m) = -\alpha m^{\beta}$, to relate average mutant fitness, W, to the number of mutations, m. Here, α expresses the rate of change in average fitness expected if mutations acted independently, and β describes the overall form of epistasis. If $\beta=1$, then mutational effects are on average multiplicative (no epistasis); if $\beta<1$, then additional mutations tend to reduce fitness less than expected from their individual effects (alleviating epistasis); and if $\beta>1$, then additional mutations reduce fitness more than expected from their individual effects (aggravating epistasis).

Sexual populations became more robust, on average, to individual mutations than did asexual ones (figure 4), with sexual organisms having significantly lower α values ($p<0.001$; table 1). The predominant form of epistasis was alleviating in both sexual and asexual organisms ($\beta<1$ based on *t*-tests, both $p<0.001$), although this directional epistasis was weaker in sexual organisms ($p<0.001$; table 1). The differences between sexual and asexual populations remained significant even when those with genome length greater than or equal to 100 were excluded ($p=0.019$ for α, $p<0.001$ for β; see electronic supplementary material).

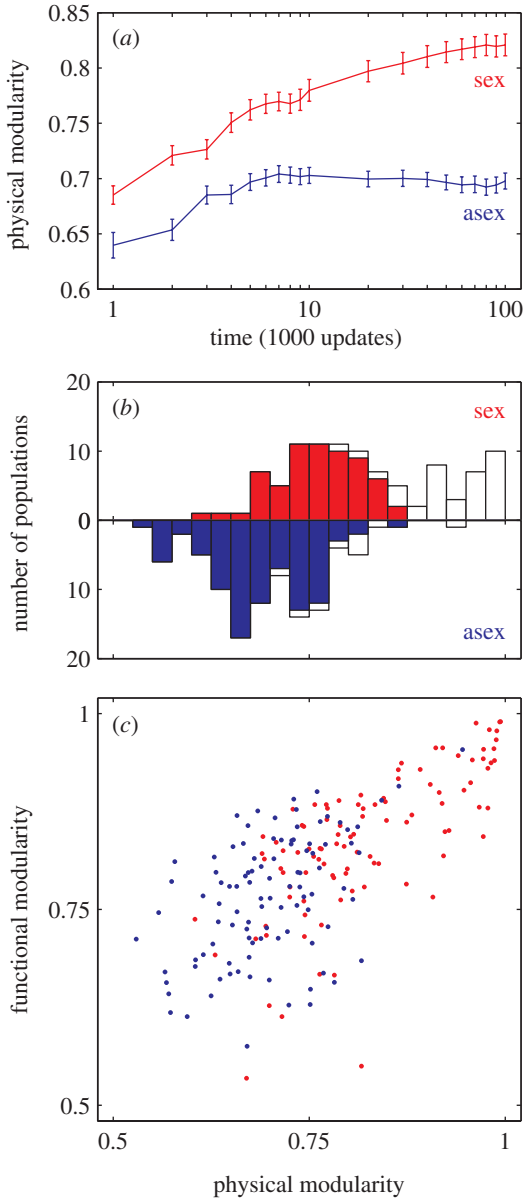

Figure 3. Genomic modularity of sexual and asexual organisms. (*a*) Trajectories of average physical modularity in evolving sexual (red) and asexual (blue) populations. Note the logarithmic time scale and that the first time-point represents 1000 updates. Error bars are standard errors of the mean. (*b*) Final distributions of physical modularity values. Open sections show populations with average genome length greater than or equal to 100 instructions. (*c*) Relationship between the final physical and functional modularity values. Each point represents a single independently evolved sexual (red) or asexual (blue) population.

4. DISCUSSION

In nature, organisms show considerable variation in their reproductive mode as well as in certain features of their genetic architecture. While there has been substantial interest in the relationship between reproductive mode and features of genetic architecture, it is difficult to establish causation (Malmberg 1977; Chao 1988; Lawrence & Roth 1996; Wagner 1996; Wagner & Altenberg 1996; de Visser *et al.* 1997; Elena & Lenski 1997; Kirschner & Gerhart 1998; Hartwell *et al.* 1999; Lenski 1999; Lenski *et al.* 1999; Ferrier & Holland 2001; Rice & Chippindale 2001; Beldade *et al.* 2002; Ravasz *et al.* 2002; Wolf *et al.* 2002;

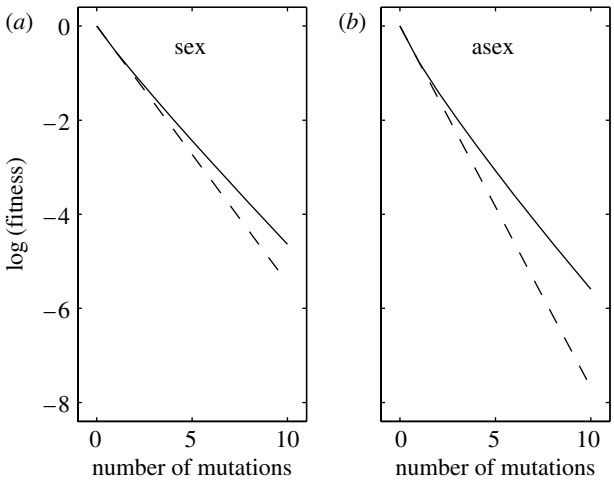

Figure 4. Average fitness as a function of the number of random point mutations in evolved (*a*) sexual and (*b*) asexual organisms. Dashed lines indicate the fitness decay functions expected under a multiplicative model of mutation interactions, using the corresponding average value for α and setting $\beta = 1$. Solid curves are the decay functions based on average values for both α and β parameters.

Pál & Hurst 2003, 2004; Wilke *et al.* 2003; Schlosser & Wagner 2004). This difficulty reflects several challenges, including a multitude of potential feedbacks between the biological features of interest, the inability to manipulate such critical features in most biological systems, and the infeasibility of performing long-term experiments to observe how manipulating one factor would impact the evolution of other features.

In this study, we have overcome these challenges by investigating the evolutionary relationship between reproductive mode and genetic architecture in digital organisms, which are self-replicating computer programs that mutate, compete and evolve. By comparing evolutionary outcomes in populations that evolved in identical environments and initially differed only in whether they reproduced asexually or sexually, we demonstrated that several features of genetic architecture were shaped by reproductive mode. The genomes of sexual organisms were significantly more modular than those of asexual organisms by two different measures (figure 3). Sexual organisms were also more robust with respect to the average effect of single mutations, while asexual organisms tended to have stronger epistatic interactions among multiple random mutations (figure 4).

An unexpected complication arose because sexual organisms often evolved much longer genomes than asexual organisms (figure 2*b*), evidently reflecting a greater propensity of the sexual populations to generate or retain genome doublings. Genome length was itself correlated with other features of genetic architecture such as modularity (figure 3*b*), which raised the question of whether differences between sexual and asexual populations in genome length might be solely responsible for the other differences in their genetic architecture. We examined this issue in two ways (see electronic supplementary material). First, we excluded from our analyses all evolved populations in which genome length had at least doubled. Second, we performed an additional set of experiments that prevented genome doublings and other large changes in genome size from occurring. In both cases, the difference in genome length became shifted in the opposite direction such that asexual organisms were in fact longer. Yet, the other differences in genetic architecture between sexual and asexual organisms retained their original form, with the genomes of sexual organisms being more modular and more robust to the fitness effects of individual mutations, while exhibiting weaker epistatic interactions among mutations.

(a) *Relationships among the genetic architectural parameters*

Previous research has demonstrated negative correlations between mutational parameters α and β using both analytical models and simulations of RNA folding (Wagner *et al.* 1998; Wilke & Adami 2001). We, too, observe a strong negative correlation between α and β in the combined dataset that includes both sexual and asexual digital organisms ($r = -0.694$, $p < 0.001$ for all 200 evolved populations and $r = -0.613$, $p < 0.001$ for the reduced dataset that excludes populations with genome length greater than or equal to 100).

We also examined the relationship between the two measures of modularity, FM and PM, and found that they are positively correlated ($r = 0.685$, $p < 0.001$ and $r = 0.491$, $p < 0.001$ for the reduced dataset). All else equal, genomes with more compact regions expressing different traits would have less overlap if those regions were randomly distributed throughout a genome, which may contribute to this association. However, the observed correlation coefficient is only moderate in magnitude, and, therefore, we conclude that these two modularity measures capture somewhat different aspects of the genetic architecture.

To examine the relationships among all the genetic architectural and performance metrics in this study, we next performed a principal component analysis (PCA) on log-transformed fitness, log-transformed genome length, PM, FM, α and β (figure 5*a*). The first two components explain over 70% of the total variance in the data. The first principal component, pc_1, largely reflects the two modularity measures, PM and FM, whereas pc_2 reflects the directional epistasis parameter, β, and fitness (figure 5*b*). Both components also show substantial loadings for genome length and mutational sensitivity, α, with these two metrics being negatively correlated, such that longer genomes tend to be less sensitive, on average, to single mutations.

We also performed a discriminant function analysis using the same six properties as in the PCA in order to examine which ones best capture the observed differences between sexual and asexual organisms. The discriminant function gives the largest and nearly equal weights to PM and β, and it would correctly classify a given organism as sexual or asexual in over 80% of all cases (data not shown). We therefore conclude that epistasis and modularity evolve differently in sexual and asexual populations and, moreover, with sufficient independence that no single feature of the genetic architecture can explain all of the differences between sexual and asexual organisms.

(b) *Implications for the evolution of sexual reproduction*

Our experiments do not specifically address the long-standing question of why sex evolved (Williams 1975;

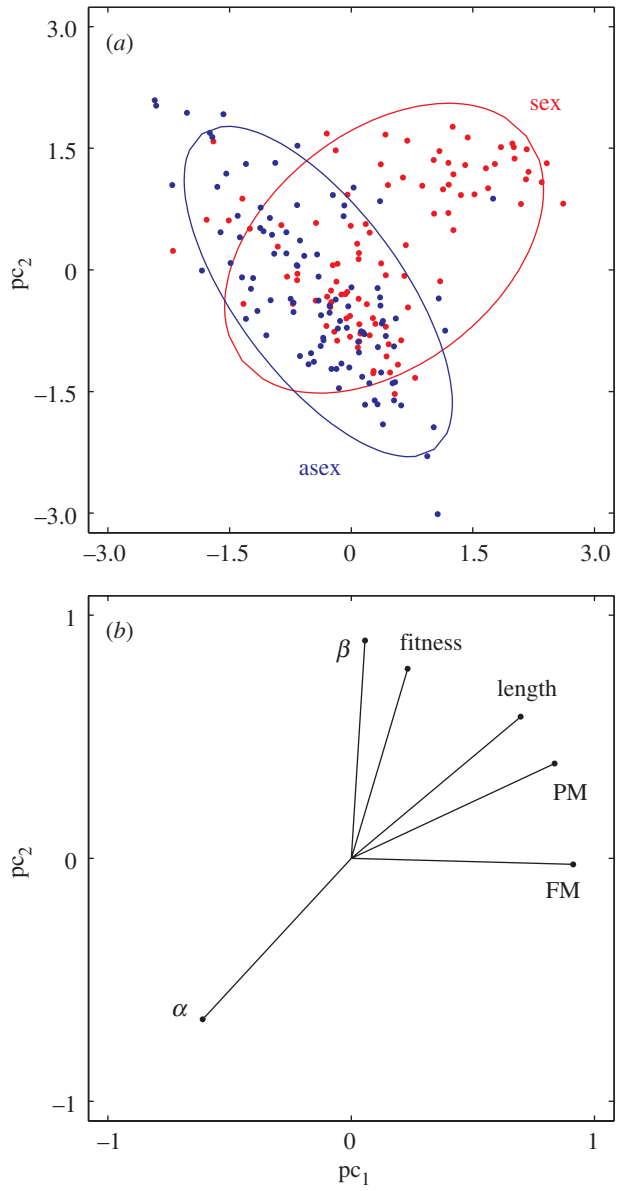

Figure 5. Principal components analysis for average fitness, genome length, α, β, PM and FM of sexual and asexual digital organisms. (*a*) Transformed data for sexual (red) and asexual (blue) populations, with 95% confidence ellipses. (*b*) Loadings for the two principal components.

Maynard Smith 1978; Bell 1982; Kondrashov 1982, 1993; Michod & Levin 1988; West *et al.* 1999). Nevertheless, some of our findings bear on this issue. In particular, our results support previous studies (Chao 1988; de Visser *et al.* 1997; Elena & Lenski 1997; Lenski *et al.* 1999; Wilke *et al.* 2003) that failed to find a preponderance of aggravating (synergistic) epistasis, which is an essential component of the mutational deterministic hypothesis (Kondrashov 1982). Importantly, we extend this conclusion by showing that an excess of alleviating epistasis exists even in sexual populations (figure 4). Our experiments also suggest two alternative hypotheses for the evolution of sex, one that is immediate and the other longer term in its consequences. First, sexual populations evolved lower values of α (table 1), indicating reduced sensitivity to the usually harmful effects of individual mutations. Thus, sexual progeny are more robust than asexual progeny owing to differences in the genetic encoding of their phenotype (e.g. canalization), as opposed to the role of recombination in purging deleterious mutations that is posited under various other hypotheses (Kondrashov 1982). In this respect, it is important to recognize that distributions of mutational effects and forms of epistatic interactions evolved freely in our experiments as populations moved across fitness landscapes, whereas these quantities are usually fixed in population genetic analyses. This finding suggests that theoreticians should give more attention to understanding the structure of fitness landscapes and how different modes of reproduction influence where populations settle on those landscapes. Second, an indirect benefit to sex could arise from higher genomic modularity, which may increase their evolvability and, thereby, promote greater fitness (Wagner & Altenberg 1996; Earl & Deem 2004). This hypothesis is more relevant to the maintenance of sexual reproduction than to its origin, however, because the benefit would not be manifest immediately (Williams 1975; Maynard Smith 1978; Bell 1982; Kondrashov 1982, 1993; Michod & Levin 1988; West *et al.* 1999). Finally, our results demonstrate that reproductive mode substantially shapes the evolution of the genetic architecture. We therefore emphasize the importance of distinguishing between the evolutionary causes and consequences of sexual reproduction.

We thank A. Kondrashov, E. Ostrowski, R. Woods and reviewers for valuable comments and discussion. This work was supported by grants DEB-9981397 and CCF-0523449 from the US National Science Foundation.

REFERENCES

Adami, C., Ofria, C. & Collier, T. C. 2000 Evolution of biological complexity. *Proc. Natl Acad. Sci. USA* **97**, 4463–4468. (doi:10.1073/pnas.97.9.4463.)

Beldade, P., Koops, K. & Brakefield, P. M. 2002 Modularity, individuality, and evo-devo in butterfly wings. *Proc. Natl Acad. Sci. USA* **99**, 14 262–14 267. (doi:10.1073/pnas.222236199.)

Bell, G. 1982 *The masterpiece of nature*. Berkeley, CA: University of California Press.

Chao, L. 1988 Evolution of sex in RNA viruses. *J. Theor. Biol.* **133**, 99–112.

Chow, S. S., Wilke, C. O., Ofria, C., Lenski, R. E. & Adami, C. 2004 Adaptive radiation from resource competition in digital organisms. *Science* **305**, 84–86. (doi:10.1126/science.1096307.)

de Visser, J. A. G. M., Hoekstra, R. F. & Van den Ende, H. 1997 Test of interaction between genetic markers that affect fitness in *Aspergillus niger*. *Evolution* **51**, 1499–1505.

Earl, D. J. & Deem, M. W. 2004 Evolvability is a selectable trait. *Proc. Natl Acad. Sci. USA* **101**, 11 531–11 536. (doi:10.1073/pnas.0404656101.)

Elena, S. F. & Lenski, R. E. 1997 Test of synergistic interactions among deleterious mutations in bacteria. *Nature* **390**, 395–398. (doi:10.1038/37108.)

Ferrier, D. E. K. & Holland, P. W. H. 2001 Ancient origin of the Hox gene cluster. *Nat. Rev. Genet.* **2**, 33–38. (doi:10.1038/35047605.)

Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. 1999 From molecular to modular cell biology. *Nature* **402**, C47–C52. (doi:10.1038/35011540.)

Kirschner, M. & Gerhart, J. 1998 Evolvability. *Proc. Natl Acad. Sci. USA* **95**, 8420–8427. (doi:10.1073/pnas.95.15.8420.)

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

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

Lawrence, J. G. & Roth, J. R. 1996 Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* **143**, 1843–1860.

Lenski, R. E. 1999 A distinction between the origin and maintenance of sex. *J. Evol. Biol.* **12**, 1034–1035. (doi:10.1046/j.1420-9101.1999.00122.x.)

Lenski, R. E., Ofria, C., Collier, T. C. & Adami, C. 1999 Genome complexity, robustness and genetic interactions in digital organisms. *Nature* **400**, 661–664. (doi:10.1038/23245.)

Lenski, R. E., Ofria, C., Pennock, R. T. & Adami, C. 2003 The evolutionary origin of complex features. *Nature* **423**, 139–144. (doi:10.1038/nature01568.)

Malmberg, R. L. 1977 Evolution of epistasis and advantage of recombination in populations of bacteriophage T4. *Genetics* **86**, 607–621.

Maynard Smith, J. 1978 *The evolution of sex*. London: Cambridge University Press.

Michod, R. E. & Levin, B. (eds) 1988 *The evolution of sex*. Sunderland, MA: Sinauer Associates.

Misevic, D., Ofria, C. & Lenski, R. E. 2004 Sexual reproduction and Muller's ratchet in digital organisms. In *Proc. Artificial Life IX* (ed. J. B. Pollack, M. Bedau, P. Husbands, T. Ikegami & R. A. Watson), pp. 340–345. Cambridge, MA: MIT Press.

Ofria, C. & Wilke, C. O. 2004 Avida: a software platform for research in computational evolutionary biology. *Artif. Life* **10**, 191–229. (doi:10.1162/106454604773563612.)

Pál, C. & Hurst, L. D. 2003 Evidence for co-evolution of gene order and recombination rate. *Nat. Genet.* **33**, 392–395.

Pál, C. & Hurst, L. D. 2004 Evidence against the selfish operon theory. *Trends Genet.* **20**, 232–234. (doi:10.1016/j.tig.2004.04.001.)

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. & Barbasi, A. L. 2002 Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555. (doi:10.1126/science.1073374.)

Rice, W. R. & Chippindale, A. K. 2001 Sexual recombination and the power of natural selection. *Science* **294**, 555–559. (doi:10.1126/science.1061380.)

Schlosser, G. & Wagner, G. P. (eds) 2004 *Modularity in development and evolution*. Chicago, IL: University of Chicago Press.

Wagner, G. P. 1996 Homologues, natural kinds and the evolution of modularity. *Am. Zool.* **36**, 36–43.

Wagner, G. P. & Altenberg, L. 1996 Complex adaptations and the evolution of evolvability. *Evolution* **50**, 967–976.

Wagner, G. P., Laubichler, M. D. & Bagheri-Chaichian, H. 1998 Genetic measurement theory of epistatic effects. *Genetica* **103**, 569–580. (doi:10.1023/A:1017088321094.)

West, S. A., Lively, C. M. & Read, A. F. 1999 A pluralist approach to sex and recombination. *J. Evol. Biol.* **12**, 1003–1012. (doi:10.1046/j.1420-9101.1999.00119.x.)

Wilke, C. O. & Adami, C. 2001 Interaction between directional epistasis and average mutational effects. *Proc. R. Soc. B* **268**, 1469–1474. (doi:10.1098/rspb.2001.1690.)

Wilke, C. O. & Adami, C. 2002 The biology of digital organisms. *Trends Ecol. Evol.* **17**, 528–532. (doi:10.1016/S0169-5347(02)02612-5.)

Wilke, C. O., Wang, J. L., Ofria, C., Lenski, R. E. & Adami, C. 2001 Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature* **412**, 331–333. (doi:10.1038/35085569.)

Wilke, C. O., Lenski, R. E. & Adami, C. 2003 Compensatory mutations cause excess of antagonistic epistasis in RNA secondary structure folding. *BMC Evol. Biol.* **3**, 3. (doi:10.1186/1471-2148-3-3.)

Williams, G. C. 1975 *Sex and evolution*. Princeton, NJ: Princeton University Press.

Winther, R. G. 2001 Varieties of modules: kinds, levels, origins, and behaviors. *J. Exp. Zool.* **291**, 116–129. (doi:10.1002/jez.1064.)

Wolf, J. B., Brodie, E. D. & Wade, M. J. (eds) 2002 *Epistasis and the evolutionary process*. Oxford, UK: Oxford University Press.

The electronic supplementary material is available at http://dx.doi.org/10.1098/rspb.2005.3338 or via http://www.journals.royalsoc.ac.uk.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.