Digital genetics: unravelling the genetic basis of evolution

Christoph Adami

Abstract | Digital genetics, or the genetics of digital organisms, is a new field of research that has become possible as a result of the remarkable power of evolution experiments that use computers. Self-replicating strands of computer code that inhabit specially prepared computers can mutate, evolve and adapt to their environment. Digital organisms make it easy to conduct repeatable, controlled experiments, which have a perfect genetic 'fossil record'. This allows researchers to address fundamental questions about the genetic basis of the evolution of complexity, genome organization, robustness and evolvability, and to test the consequences of mutations, including their interaction and recombination, on the fate of populations and lineages.

Genetic algorithm
A computational method that uses Darwinian methods to search for a rare solution (encoded into a symbolic string) within a large search space. Typically, the best strings in a population are mutated and recombined to form the next generation, whereas inferior strings are removed.

Keck Graduate Institute of Applied Life Sciences, 535 Watson Drive, Claremont, California 91711 and Digital Life Laboratory 136-93, California Institute of Technology, California 91125, USA.
e-mail: adami@kgi.edu
doi:10.1038/nrg1771

Although evolutionary biology has been making steady progress since Darwin's time, we are still far from reaching a satisfactory understanding of the genetic bases of evolutionary change and adaptation. The recent emergence of quantitative evolution experiments using microorganisms[1] has moved the field closer to the reciprocity between experiment and theory that underlies such successful disciplines as physics and chemistry. However, a significant gap remains, because long-term evolution experiments in complex environments require unreasonable investments of time and resources.

More recently, digital organisms have been increasingly used to address some of the fundamental problems in evolutionary biology[2,3] (BOX 1). After all, Darwin's principles of evolutionary change are universal: they do not refer directly to the particular carriers of information, or to any particular genetic mechanism[4,5]. Rather, they apply to any system in which information is stored, inherited with variation, and determines the differential survival of the carrier. Therefore, Darwinian systems should also be realizable within a computational chemistry. In a computational environment, information can be copied (for inheritance) with different degrees of variation (through errors), and the information carried by computer programs can be expressed through the execution of the program. Importantly, such an implementation of Darwinian mechanisms is inherently different from a simulation (such as a genetic algorithm), because the fitness of a program is not determined *a priori* by the user. Instead, as for biochemical life, those lineages that survive the competition for space, time and resources are the most fit — in hindsight.

Experiments can be designed for digital organisms that are either impossible or impractical when using 'biochemicals', and can uncover new principles that might open up new avenues of research. For experimental evolutionary biologists, digital organisms represent the realization of a dream they have pursued ardently and imaginatively: to be able to carry out repeatable, statistically powerful experiments under fully controllable conditions, with a perfect fossil record. Although digital genetics excels in these attributes, it has important limitations (BOX 2), and many questions concerning the causes and consequences of evolution still require conventional evolutionary genetics. However, some exciting and fundamental problems in evolutionary biology can be addressed using digital organisms, and the scope of digital genetics is likely to expand.

Here I give an overview of digital life systems in general and the Avida platform (the most widely used system for digital genetics) in particular. I then examine some key areas of evolutionary genetics in which digital organisms have already begun to provide important insights.

Digital life

Digital life stems from the realization that self-mutating computer viruses could be studied within simulated, rather than real, computers. Because any computer can be simulated within any other sufficiently complicated one, it is possible to create, within a specifically prepared part of a standard computer, an artificial Petri dish where computer programs can self-replicate and compete for survival. Because the space in this simulated world is limited and the replication process can be made to be

Box 1 | **The 'big questions' in evolutionary biology**

Can evolution be predicted? How much do chance events shape the outcome of evolution? What is the role of the history of a lineage and of adaptation? What would happen if we 're-ran' evolutionary history[74–78]?

How does speciation occur? The original question studied by Darwin still has no complete answer. Can organisms speciate in sympatry? What factors contribute to species diversity and which mechanisms impede it[70,79,80]?

What is the advantage of sex? Under which circumstances is recombination beneficial to the lineage? How can such benefits overcome the twofold cost of sex[64,65,81–83]?

Is there a trend in the evolution of biological complexity? Does Darwinian evolution imply ever-increasing complexity[35,40,84]? If there is such a trend, what role does co-evolution have? What factors are responsible for complexity crashes[40,41,85]?

How did life and evolution begin*[86,87]*? How do chemical systems change from a purely thermodynamic regime to an information era of evolutionary replication[88,89]? Are there general principles involved in this transition[90,91]? What is the probability of life in a non-terrestrial chemistry?

Of the five big questions listed above, the first four are being addressed using theoretical, computational and experimental approaches, including digital genetics. For the fifth (How did life begin?), experiments with biological organisms are obviously impossible, but there is a chance that digital genetics can shed light on some aspects of this question in the future, even though self-replicators are also extremely rare in the digital chemistry (estimated at 1 self-replicator among 10^{15} randomly generated sequences at the most). Recently, a 'biosignature' algorithm that was originally designed to detect extraterrestrial life successfully 'detected' digital life (E.D. Dorn and C.A., unpublished observations).

Sympatry
The condition in which the distributions of two species overlap and hybridization between taxa would be possible if they were not reproductively isolated by factors other than spatial separation.

Robustness
The ability of an organism's genome to withstand point mutations (substitutions) without losing fitness. Numerically, this is expressed as the fraction of all possible single substitutions that do not change the organism's fitness. At high mutation rates, an increase in mutational robustness is selected for.

Central processing unit
(CPU). The 'brain' of a computer, where the information is processed. In Avida, a simulated CPU translates the genetic information into actions that represent the phenotype of the information.

Carrying capacity
In ecology, the maximum number of organisms of a particular species an ecosystem can support. In the single-niche or multi-niche digital system Avida, this is simply the total number of organisms, which can be set by the user.

inaccurate, the three basic ingredients of Darwinian evolution are present — inheritance, variation and selection. The adaptive complexity that arose from the first such implementation, Tierra[6], confirmed the enormous potential of digital life.

Here I describe the basic mechanics, genetics and metabolism of 'avidians'[7–9] because they are the most frequently used digital organisms in evolutionary research, and because their main design features are common to most forms of digital life. Avidians are computer programs that have been written in a special-purpose language that can only be interpreted within the simulated computer that is created with the Avida software. These programs encode the ability to self-replicate and carry out computations, and segments of the code that are responsible for a particular functionality are referred to as the organism's genes. Each gene is made up of sets of instructions, each of which carries out a particular function. The language usually has between 20 and 30 instructions[10] and is quite robust to mutations: depending on the program, between 20% and 70% of single-instruction substitutions do not affect the program's function. A similar range for the neutrality of a sequence has been observed for proteins in biological organisms[11,12].

Replication. Avidians replicate by executing their code, copying it, instruction by instruction, into fresh memory that is obtained by elongating their memory space, and then dividing off the copy. The particular sequence of instructions in the code represents the organism's genotype. The code itself is circular and is executed by an instruction pointer that roams the sequence, much like a polymerase does when transcribing bacterial DNA (FIG. 1). The execution of an instruction affects the state of the central processing unit (CPU) that is attached to each program. This CPU (depicted schematically in FIG. 1a) represents the state of the organism, and is therefore the means of expression of the information, much like the sum of expressed proteins and their states represents a cell's state. The sequence of changes that are the result of code execution define an organism's phenotype.

When a program has successfully copied its code and divided, the Avida software places the offspring into the population. The user can select how this is done, but usually the offspring is placed either next to the progenitor on a two-dimensional regular grid (emulating growth on a Petri dish), or randomly in a well-mixed population without spatial structure (simulating the growth dynamics of cultures in a shaken vial or chemostat). In either case, a random member of the population is removed to make room for the new arrival. The population size is usually kept strictly constant to enforce selection, but it is possible to implement variable population sizes (up to the carrying capacity) by introducing other random death events.

Mutations. Digital offspring can differ genetically from their parents for several reasons. The most common cause is a single substitution error during the copying process, which is due to an inherent inaccuracy in the copy instruction. The probability of error per instruction copied can be set by the user, and the number of errors in a fixed time interval is given by a Poisson distribution. Besides these copy mutations, the user can select a rate of insertion and deletion mutations, and a probability of genetic defects occurring on code separation (division). It is also possible to cause substitutions that affect every existing individual, irrespective of whether the organism is in the process of copying or not. This type of mutation mimics the effect of cosmic rays or mutagenic chemicals.

Apart from these explicit mutations, genetic differences can also arise through implicit mutations; for example, if the division between parent and offspring code is asymmetrical. Implicit mutations usually have large effects, and are often lethal. Another type of implicit mutation can occur if the replication loop is sloppy such that the parent overwrites parts of its own code rather than copying into the pristine space of the offspring. Finally, genetic changes can be introduced through 'code recombination', at a rate that is set by the experimenter. This is initiated by a special divide instruction, and results in the interchange of genetic material with a nearby organism, in a form of digital sex. Most experiments have so far been run in the asexual mode.

Selection and population dynamics. As for microbial and viral self-replicators, the fitness of a digital organism is given by the growth rate of the clone it gives rise to. This rate is a function of two factors: the efficiency of the copy procedure, and the ability to exploit the environment to speed up replication. As with biochemical organisms, the replication of digital information costs energy. For avidians, energy is dispensed in units that allow the execution of a single instruction — a so-called 'single-instruction processing' unit (SIP). There are two mechanisms to receive SIPs: one is passive and provides a basic amount

Box 2 | **Limitations of digital genetics**

Digital organisms have obvious advantages for the study of evolutionary and population genetics, owing to the organism's short generation time and the ease of preparing perfectly known environments and controls. Functional genetic studies in digital organisms, however, are limited to abstract investigations of how information coding is influenced by different selective pressures and mutational mechanisms. Because digital genomes are expressed directly through code execution, topics that involve the molecular mechanisms of transcription, translation and intron splicing, for example, are beyond the scope of digital genetics. Digital organisms have only limited means of expression regulation, and there is no developmental phase. Moreover, epigenetic modifications are absent, and although code recombination after replication is an option, digital organisms undergo neither syngamy (the fusion of two genomes) nor meiosis. Finally, it should always be kept in mind that digital genomes are extremely short compared with even the smallest free-living organisms and most viruses. Consequently, although the mutation rate per genome is usually comparable to those of most RNA viruses[92], the rate per gene is unusually high.

of energy that is proportional to sequence length, the other is active in that an organism has to earn it. The mechanism that allows organisms to gain energy actively is logically similar to a metabolic pathway in which molecules are converted enzymatically to form products, releasing ATP. Digital organisms earn energy by evolving computational pathways that convert input numbers into output numbers, and each of these calculations releases a specific number of SIPs (BOX 3). The speed of code execution, and therefore the growth rate of a clone, depends directly on the SIPs earned by an organism during its life cycle. Note that although the total amount of SIPs earned by the population per update can increase, the total amount of CPU time given to the population is constant. Therefore, SIPs only affect the relative rate of replication.

With inheritance through replication, variation through mutations, and selection that is due to finite resources of space and time, Darwinian population dynamics is ensured. At low mutation rates, periodic selection drives beneficial mutations to fixation at a rate that is proportional to their fitness advantage. Each mutation is fixed before the next one arises, so successive dominant genotypes replace each other[13,14]. However, at high mutation rates clonal interference[15] is observed, and population dynamics conform to a quasispecies model in which mutational robustness has a key role in survival[16] (see below).

Tools for analysing digital organisms

Several powerful tools are available for the analysis of digital populations. I focus here on two that are relevant to the studies described in this review.

Functional genomic array. Similar to DNA-based code, the function of a digital genome is hard to decipher from the code alone (although this is not impossible for a trained code-writer). Instead, a computational assay is used (BOX 4). Instructions are knocked out one after the other (by replacing them with an inert instruction), and the resulting functionality and fitness are evaluated by executing the organism outside the population and recording the number of offspring that are generated per unit of time, as well as the computations that are carried out. This tool is the digital equivalent of a classical genetics knockout experiment, except that the functional components of single genes can be dissected (rather than just the effect of the presence or absence of a gene) by studying the effect of all possible mutations within each gene.

Phylogeny tool. Phylogeny reconstruction is the central element of evolutionary genetics. For digital organisms, all descent information can be stored, but because each experiment typically yields millions of genotypes, it is not practical to retain it for all types that ever existed. Instead, descent information is only kept for those types that are present in the population at the most recent point in time. Once a lineage becomes extinct, that branch of the tree is eliminated from the memory. The retained information is sufficient to create a phylogenetic-depth tree that reflects the dynamics within the adapting population, by plotting the number of individuals at depth d that have arisen from the founding ancestor, against time (FIG. 2). The line of descent of any genotype present in the population when the experiment was stopped can be superimposed onto the phylogenetic-depth tree. This phylogeny tool, together with the use of functional genomic arrays, has been invaluable in addressing the questions that are outlined below using digital genetics.

Mutations and epistasis

Although classical genetics treats the effects of mutations mainly as independent, recent evidence implies that many mutations interact with each other (a phenomenon that is termed epistasis), either within or between genes[17]. Furthermore, the form of the interaction — whether a deleterious mutation at one locus increases (synergistic epistasis) or decreases (antagonistic epistasis) the deleterious effect at another — has been shown to have an important role in evolution. According to Kondrashov's mutational deterministic hypothesis[18,19], the average direction of epistasis (synergistic or antagonistic) between mutations has an important — perhaps determinant — role in an organism's capacity to purge mutations and maintain a minimal mutational load. But the experimental determination of the average direction of epistasis in biological organisms is difficult and time-consuming owing to the enormous number of combinations of mutants that must be studied[20]. Therefore, there is no consensus about the direction of interactions — for single genes or whole genomes — in any organism (but see REF. 21). Digital genetics allows us to study the direction of epistasis with high statistical accuracy, and its influence on other determinants of evolutionary fate.

Lenski *et al.*[22] systematically studied the average direction of epistasis in 87 digital organisms that were evolved independently from the same simple progenitor. A control group of simple genomes was obtained by re-adapting ('de-evolving') each of these populations in an environment that favoured short sequences. The direction of epistasis was determined for the complex and de-evolved sets by measuring the effect of all possible one-, two- and three-point mutations (as well as a sample of higher multiples of up to ten-point

Poisson distribution
The distribution of the number of occurrences of a discrete number of events during a fixed amount of time. For an average rate of occurrences λ, the probability of observing k events during time τ is $P(k) = e^{-\lambda\tau}(\lambda\tau)^k/k!$

Implicit mutation
A genetic change that appears as a consequence of a faulty replication process rather than owing to an explicit mutation agent such as substitution, insertion, deletion or recombination. Examples include repetitions, as well as excisions of whole segments of code.

Replication loop
In digital organisms, the segment of code responsible for the duplication of genetic information. In most cases, this segment is 'looped over' many times to effect replication.

Update
An arbitrary unit of time in digital life experiments, during which every member of the population executes a finite number of instructions (usually set as 30). The number of updates that elapse during one generation is not fixed because the time to produce an offspring changes during evolution.

Fixation
In population genetics, the fixation of an allele or trait is defined as the moment at which every member of the population carries that allele.

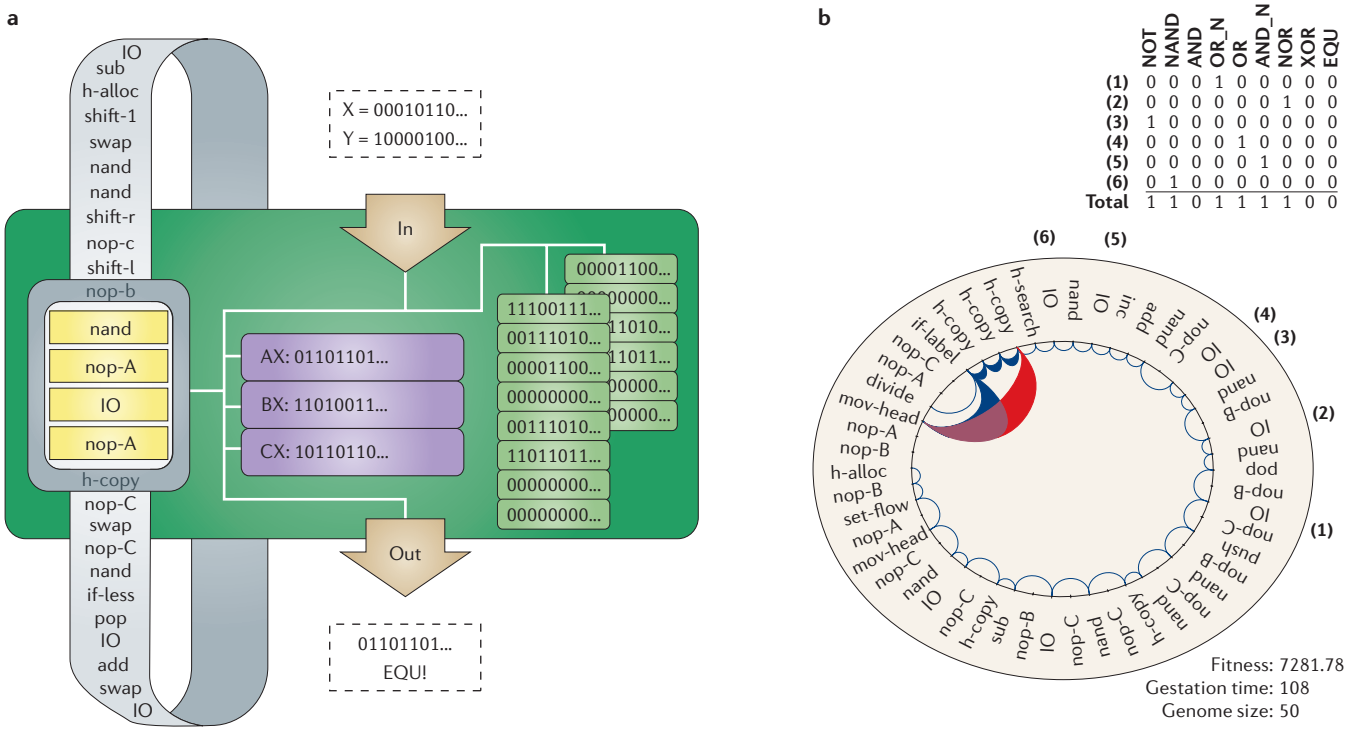


Figure 1 | **Avidian genomes. a** | A digital organism and its central processing unit (CPU) in Avida. The virtual CPU has three registers (AX, BX and CX, shown in purple), as well as input and output buffers (boxes with dashed outline) and a double stack (light green). Registers, buffers and stacks can hold numbers that are used by the organism both for replication and for computation — that is, their metabolism. The genome of a digital (shown in grey) consists of the computational instructions that make up its circular code, within which groups of instructions that carry out a specific function comprise genes. The code is executed serially (see part **b**), and affects the movement of numbers between registers, stacks and input/output buffers. For example, the first part of the code snippet highlighted in yellow carries out a logical 'nand' operation on the numbers contained in the BX and CX registers and places the result in the AX register (nand and nop-A). The pair IO and nop-A in turn writes the result of the AX register to the output. The CPU also communicates with the environment. For example, on inserting the instruction 'divide', the CPU signals that the code needs to be separated and placed into the population. **b** | Execution of an avidian genome. A trace of the execution pattern of the instruction pointer is shown. The pointer moves along the sequence of instructions, thereby executing the computational genes. Anticlockwise movement is indicated in blue, clockwise movement in red. The phenotypic characteristics of the organism — in the form of the computational tasks that are carried out — are listed in the table. The genome uses the computational instructions to carry out, in this particular organism, six of the nine computational tasks that are listed in the table (three are not performed: 'AND', 'XOR' and 'EQU'). Therefore, it can be said to carry these six genes. The numbers in the outer ring indicate the IO (input/output, that is, read/write) instruction that triggers the reward associated with the computational task. Therefore, the instruction 'IO' that is marked as (6) triggers the reward for the 'NAND' gene. The completion of this task is rewarded by extra CPU time (in the form of SIPs (BOX 3)) for the organism, thereby increasing its speed of replication. Part **a** reproduced with permission from *Nature* REF. 44 © (2003) Macmillan Publishers Ltd. Part **b** is courtesy of D. Misevic.

Clonal interference
The competition between beneficial mutations or alleles before fixation. In clonal populations within a single niche, only one of several competing (interfering) mutations can go to fixation.

Quasispecies model
A theoretical description of a population of self-replicators at high mutation rate, which is characterized by a 'cloud' of mutationally interdependent types rather than a single, dominant wild type. 'Species' here refers to the species concept in chemistry rather than biology.

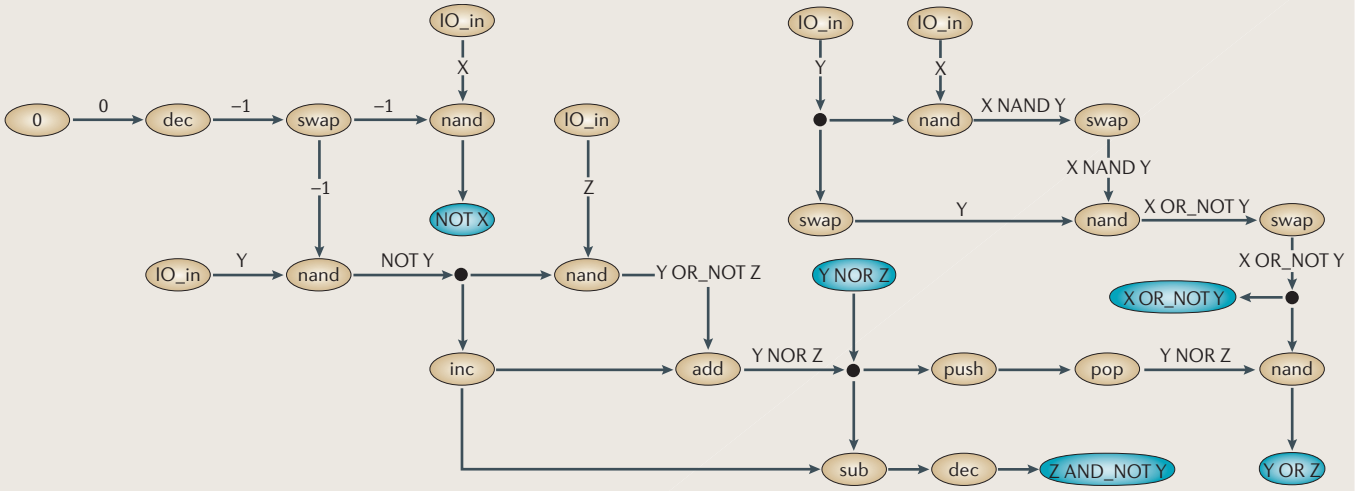
Phylogenetic depth
The total number of generations in which an offspring organism has differed from its parent, or the cumulative number of genetic changes that separate an organism from the ancestor of the lineage.

Kondrashov's mutational deterministic hypothesis
The hypothesis that sex will evolve and be maintained in populations at high mutation rates if mutations interact in a synergistic (that is, aggravating) fashion, but not if mutations interact antagonistically.

mutations) on the fitness of every organism in the set. In this manner, Lenski *et al.* determined the 'fitness decay function', $w(n)$, of each parent genome as a function of the number of mutations, n, by fitting the function: $w(n) = w(0)\exp(-\alpha n^{\beta})$.

The parameters α and β reflect the average effect of single mutations (the mutational robustness) and the average direction of epistasis, respectively, on the mean fitness w. If mutations are independent on average, we would find that $\beta = 1$. Synergistic epistasis between mutations, that is, interactions that reinforce the deleterious effects of multiple mutations, results in a coefficient $\beta > 1$, whereas antagonistic (mitigating) interactions lead to $\beta < 1$. Lenski *et al.* found that $\beta < 1$ for most of these organisms, but with a significant variation that is due to differences in genome complexity. Epistasis was significantly antagonistic for organisms that have complex genomes, whereas a multiplicative model (whereby $\beta = 1$) was sufficient to explain the fitness decay of the simple genomes, which had shrunk to retain only their replicative gene. Therefore, the simple genomes were more sensitive to single mutations, whereas the complex set showed robustness to both single and multiple mutations through antagonistic epistasis. However, the lack of epistasis for the simple genomes in general does not mean that pairs of mutations did not interact. Instead, both the simple and the complex set showed either antagonistic or synergistic effects for about three-quarters of all pairs, but these all but cancelled out when they were averaged for the simple set.

Box 3 | **Digital metabolism**



The digital world uses a computational chemistry, where instead of creating product R from compounds A and B, a calculation converts random binary numbers — for example, *X* and *Y* — to a result $R_i(X,Y)$. This is dependent on the correct sequence of instructions to carry out this calculation having evolved in the digital genome. In Avida, up to 68 possible calculations can be rewarded, with the rewards being in the form of single-instruction processing units (SIPs — the computational equivalent of ATP). These calculations comprise all possible logical operations that can be carried out using one to three binary inputs (there are 2 one-input reactions, 8 two-input reactions and 58 computations on three inputs). These computational reactions differ in complexity because the genetic language includes only a single logical instruction, nand, that can be used to produce different results. For example, the result 'NOT X' is obtained from the input 'X' through a single nand, whereas 'X OR_NOT Y' requires two nands.

In the above pathway (which is automatically generated from an avidian genome using a tool developed by Weise (D. Weise, personal communication)), ovals marked 'IO_in' denote the uptake of a number from the environment. These numbers are then processed in parallel by separate computational instructions (yellow ovals) that together constitute genes, giving rise to the logical outputs (shown next to the arrows). The final rewarded logical outputs ('NOT X', 'Y NOR Z' and 'Z AND_NOT Y' in the left part of the pathway and 'X OR_NOT Y' in the right part) are shown as cyan ovals and are rewarded with SIPs. The most complex rewarded calculation in this pathway is 'Y OR Z', and is formed from the other rewarded outputs (more complex calculations trigger more SIPs). As organisms acquire the genes that contain the instructions to carry out more and more complex operations, a type of logical metabolism develops where different genes carry out parts of calculations that are then picked up by other genes.

For all 174 genomes, this study showed a strong correlation between α and β, implying that these parameters do not change independently[23]. Instead, the evolution of robustness to single mutations (smaller α) goes hand-in-hand with less antagonistic interactions (where β is closer to 1). This intriguing result indicates that there might be an evolutionary path towards synergistic interactions (which favour recombination in Kondrashov's theory), if genomes are forced to develop a small α. However, it is unclear how the threshold $\beta = 1$ can be breached for asexual genomes, as discussed in more detail below.

Many uses of digital organisms in understanding epistasis can be predicted. For example, San Juan *et al.*[24] have shown that antagonistic interactions of deleterious mutations in vesicular stomatitis virus are accompanied by antagonistic interactions between beneficial mutations — something that can be verified in digital organisms. The interaction of deleterious mutations with beneficial ones ('decompensatory pairs')[17] is even less understood, but might be important in the emergence and maintenance of sexual recombination[25].

Mutational load
The fitness reduction of a population owing to mutations in the gene pool.

Genetics of mutational robustness

The complexity of cellular-repair mechanisms (such as proofreading and error correction) highlights the importance of reducing the mutational load that is brought about by the noise inherent in genetic information processing. Digital genetics has revealed that even in the absence of explicit repair mechanisms there are considerable differences in how robustly genomes respond to mutations (reflected by differences in the average effect of single mutations, α)[22]. But this analysis did not shed light on whether these differences are adaptive. It is also not clear how these differences arise genetically, that is, how two functionally similar or even identical genomes can encode the same information in a more or less robust manner.

To study the evolution of mutational robustness in digital organisms, Wilke *et al.* created pairs of sequences with disparate robustness (differing α) and observed the effect on the long-term survival of the resulting populations[16]. The pairs of sequences were obtained by subjecting evolved sequences (from the 'complex' set discussed above) to environments that had

Box 4 | **The functional genomic array of an avidian**

The functional array shows the effect of knocking out individual computational instructions on the functions (or genes) of a digital organism. The 60 instructions of the organism are displayed in the left-hand column along with their alphabetical and mnemonic codes. For this experiment, the instruction set consisted of 26 unique instructions. The other 9 columns are the functions that were assayed. Besides the activity of the replicative gene 'REPL' (first column of functions), the authors studied the activity of the computational genes NOT, NAND, AND, OR_N, OR, AND_N, NOR, XOR, and EQU (subsequent columns). The wild-type organism could carry out the six instructions illustrated in green and could not carry out those illustrated in red. The assay looked for a loss-of-function in the former and a gain-of-function in the latter. The effect on each function of knocking out each instruction is shown: white indicates that the function is unaffected, red means that the function is turned off, whereas green signals that the function is turned on instead. For example, knocking out the highlighted instruction ('46 g push') activated the AND gene and inactivated the EQU gene[44]. Because organisms can carry out each logical calculation several times, this array can also distinguish in principle between increased versus decreased activity. The tool allows for a precise determination of gene location, whether and where the functions overlap, as well as whether gene functions are linked to each other. Figure reproduced with permission from *Nature* REF. 44 © (2003) Macmillan Publishers Ltd.

Instruction	REPL	NOT	NAND	AND	OR_N	OR	AND_N	NOR	XOR	EQU
1 r h-alloc	■									
2 m dec						■	■	■		■
3 z set-flow	■									
4 a nop-A	■									
5 v mov-head	■									
6 c nop-C	■	■			■	■	■	■		■
7 g push										
8 m dec										
9 c nop-C							■			
10 i swap										
11 q IO										■
12 q IO										■
13 p nand		■								
14 t h-copy										■
15 q IO		■		■		■		■		■
16 p nand				■		■		■		■
17 q IO				■		■	■	■		■
18 c nop-C						■	■	■		■
19 p nand						■	■	■		■
20 c nop-C				■		■	■	■		■
21 t h-copy										■
22 l inc						■		■		■
23 e if-less										
24 t h-copy										■
25 n add						■	■	■		■
26 c nop-C						■	■	■		■
27 o sub							■			
28 g push				■		■				■
29 c nop-C						■				■
30 b nop-B										■
31 e if-less										
32 a nop-A										
33 m dec							■			
34 q IO					■		■			■
35 d if-n-equ										
36 t h-copy										■
37 q IO					■	■		■		■
38 c nop-C					■			■		■
39 p nand			■		■	■				■
40 t h-copy										■
41 i swap						■				■
42 p nand					■					
43 q IO					■					■
44 f pop						■				■
45 p nand						■				■
46 g push				■						■
47 q IO						■				■
48 x get-head										
49 u h-search	■		■	■						■
50 t h-copy										
51 y if-label										
52 c nop-C										
53 u h-search										■
54 a nop-A										■
55 s h-divide	■									
56 t h-copy										
57 t h-copy										
58 t h-copy										
59 v mov-head	■		■							■
60 a nop-A										
State changes	8	3	3	7	7	19	12	13	0	35

markedly different mutation rates (2 versus 0.5 mutations on average per sequence and generation), under the assumption that populations that are subjected to the high rate will recode information in a robust manner[26]. This assumption was correct, but the robust sequences paid a price by forgoing replication speed, putting into question the adaptive value of mutational robustness.

Wilke *et al.* asked whether the deficit in replication speed can be compensated by mutational robustness, and how this compensation depends on the mutational environment. Classical population genetics does not allow for such a circumstance because it assumes that a genome's response to mutation is determined entirely by the rate of mutation (see REF. 27 for a full mathematical explanation). In these experiments, pairs of organisms, in which one had a lower replication rate but higher mutational robustness than the other, were put into direct competition at differing mutation rates. For low rates, classical population genetics correctly predicts the outcome: the faster replicator drives the slower one to extinction regardless of a difference in mutational robustness. But for each of the pairs, there was a critical mutation rate at which the outcome of the competition flipped: at high mutation rates, the more robust genomes reliably replaced the faster replicators. Indeed, this change of fortunes is predicted by an extension of classical population genetics theory that describes the situation at high mutation rates, which is known as the quasispecies theory[28–32]. Therefore, mutational robustness is an evolvable genetic trait[33,34].

Evolution of complex genes and genomes

The study of the evolution of complex features is at the intersection of evolutionary and functional genetics. Which genetic changes occurred when, how were they transmitted, and what functional advantage (if any) did they confer? Although the fossil record that documents the evolution of complex features is sparse, the evidence that we do possess is consistent with a Darwinian sequence of events that builds complexity, mutation by mutation, mostly through gradual changes that create complex genes, but sometimes taking advantage of radical genome reorganizations, in the background of an ever changing environment and adapting to ever-changing targets[35]. Although such a (mostly) gradual mechanism is sometimes anecdotically questioned, it was anticipated by Darwin[36], and there is now ample experimental evidence that even complex organs such as eyes can evolve in such a manner[37–39].

Digital organisms are ideal for testing how complex genes evolve, as they allow us to ask by which path and using what kind of genetic changes a particularly complex (digital) gene emerges, how likely the evolutionary path is, and how repeatable it is. But defining the concept of complexity in biology has been controversial[40–42]. The intuitive idea of structural complexity (the number of different parts and connections) is difficult to capture, as is functional complexity[43] (the number of different functions an organism can carry out).

However, because we can simply count all the functions that digital organisms can carry out, functional complexity is straightforward in digital organisms. Taking advantage of this, Lenski *et al.* studied the evolution of complexity by tracking every mutation on the line of descent of a complex gene[44]. They studied a

digital gene that carries out the logical EQUALS function (a function that compares two binary numbers and returns a 0 for each bit that differs, and a 1 for each bit that is the same). Such genes (which require a minimum of 19 instructions, but usually take more than this) evolved in 23 out of 50 independent experiments that were seeded with a sequence that could only carry out replication. Although extending evolutionary time tenfold increased the number of successful experiments to 44, a statistical analysis indicated that the number was unlikely to increase further with extra time (C. Ofria, personal communication).

For each experiment, the complete line of descent for each member of the population was reconstructed using the phylogeny tool (for an example see FIG. 2), and every mutation was characterized. Owing to the relatively high rate of mutation ($\mu = 0.225$ per genome per generation for the ancestral type, and higher as the sequence length increased), various changes were observed. Single-point mutations were dominant, but double and even triple mutants were not uncommon, along with simple insertion and deletion mutations and (more rarely) complicated implicit mutations that affected the sequence as a whole. More importantly, not all the changes were beneficial. Along with neutral mutations, there were several slightly to drastically deleterious mutations on the line of descent. Although slightly deleterious mutations are expected, the drastically deleterious ones were surprising because to appear on the line of descent, an organism has to live long enough to give rise to at least one offspring. These mutations that had large detrimental effects were followed relatively quickly by mutations that 'rescued' the injured genotype from extinction. Often, the deleterious mutation was necessary for the fitness effect of the subsequent mutation. Evidently, the evolution of complex features does not necessarily take a straight path upwards, but takes occasional detours that lead through low fitness territory.

The appearance of strongly deleterious mutations on the line of descent cannot simply be explained by the possibility of reversion owing to the clear evidence of epistatic interactions between mutations. Indeed, standard population-genetic arguments would have a difficult time explaining the statistical prevalence of strongly deleterious mutations that reached fixation because genetic drift is too weak at these population sizes. However, recent modifications to the theory can explain these observations[31,45] because they take into account changes in the genetic background that are brought about by mutations. In particular, deleterious mutants that are part of the quasispecies of genotypes that surround the wild type are 'strengthened' by the presence of neutral and advantageous mutants nearby[31].

Selective pressures on genome organization

It has long been recognized that the genome is not a passive participant in biological evolution, but that it has a role beyond being the central repository of information[46–48]. Natural selection must have shaped genome architecture to maximize the probability of long-term success of future generations[49]. However, the means to achieve this goal are as varied as the environments that genomes find themselves in.

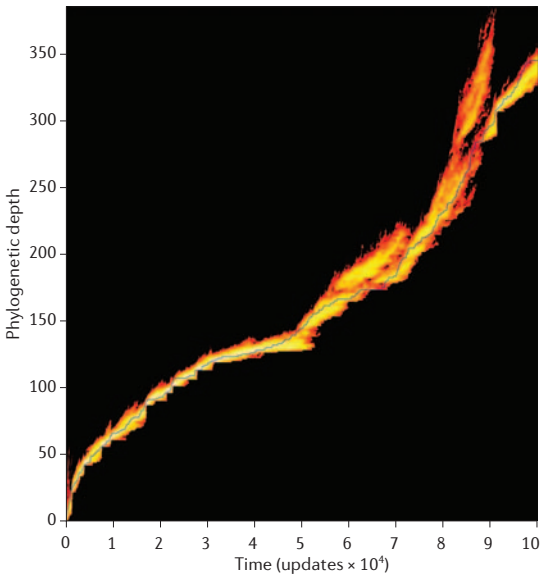


Figure 2 | **Use of a phylogeny tool for reconstructing the evolution of digital organisms.** Phylogenetic depth of a population of digital organisms that evolved the EQU gene before update number 30,000 at a depth of 111 (REF. 44), with the line of descent of the most abundant organism after 100,000 updates superimposed in blue. Different orange hues indicate the relative abundance of genotypes present in the population (yellow depicts the most abundant). The line of descent usually follows the depth of the most abundant genotype, but with significant exceptions: sometimes an organism on the direct line of descent was never represented by more than one individual. Different slopes indicate different speeds of evolution. Because of the competitive exclusion principle[93], only one lineage can ultimately survive in a one-niche experiment. In multi-niche experiments[69,70], different branches coexist stably. Reproduced with permission from *Nature* REF. 44 © (2003) Macmillan Publishers Ltd.

We have seen previously that high mutation rates favour genomes that mitigate the effect of deleterious mutations. One way to achieve this is to decrease the amount of epistasis between genes or mutations, so that single mutations do not affect more than a single locus or gene[50]. Conversely, low mutation rates, combined with the necessity to pack as much information into a limited amount of space, exert the opposite pressure. This results in overlapping genes, where mutations are more likely to have deleterious effects. Overlapping genes occur mostly in the small genomes of viruses and parasites[51,52], but are also common in bacteria[53]. However, the pressure to compact information into limited space is not the only explanation for the frequency of overlapping genes, nor is it immediately clear that it necessarily leads to a selective advantage[54]. For example, in bacteria it seems that the overlap has a mostly regulatory function[53].

Testing an hypothesis about the selective advantage of a coding strategy is notoriously difficult because it requires controlled evolution experiments using similar organisms that have different genetic architectures.

Genetic drift
In evolutionary genetics, a random process that can lead to the fixation of neutral and even deleterious alleles.

Although there are no reading frames for digital organisms, gene overlap can be studied by allowing multiple expression from multiple instruction pointers that roam the genome, which is akin to overlapping in-phase reading frames. Ofria and Adami[55] tested the degree to which multiple expression was acquired in populations that adapt to a complex fitness landscape, and compared the overall dynamics with controls that were restricted to a single-instruction pointer. In these experiments organisms can acquire multiple expression by incorporating a sequence of instructions that creates an extra execution thread, akin to the evolution of an extra promoter element. Such multiple expression evolved consistently (because the extra functional capability is immediately beneficial), but overlapping genes proved to be a burden in the long run. In the presence of multiple expression, when sequence length was unconstrained, the mean fitness increase was slower, the sequence length was significantly shorter, and the average genome neutrality (the probability that a mutation does not affect the replication rate) was significantly less than in the controls[55]. Therefore, we expect that when there is direct competition, the short-term benefit of added functionality would ultimately prove detrimental owing to the impaired evolvability of overlapping genes. This has also been observed in the overlapping genetic code of the hepatitis B virus[56]. The digital experiments did not test whether multiple expression was advantageous when genome length was limited, and this might be the winning strategy under such circumstances.

If multiple expression and gene overlap impairs evolvability, what factors promote it? Gene modularity has been suggested, as it can be seen as the antipode of overlap[46]. Mitigating gene overlap by recoding strongly epistatic genes in an independent manner has been shown to have a significant role in the evolution of mutational robustness in digital organisms[50].

Sex in digital organisms

Digital organisms represent a unique opportunity to study the origin and maintenance of sexual replication because both the mode of replication and the factors believed to affect it can be controlled. However, care must be taken because digital sex is different in important details from biochemical sex. For example, digital organisms do not suffer from the twofold cost of sex, and are therefore much closer to the genetically engineered forms of yeast that were used by Goddard *et al.*[57] than to organisms that have a defined sex. These authors studied the effect of recombination on evolution with yeast strains that were engineered to be either sexual or asexual, in an experiment that used similar strategies to those of digital genetics. Recombining strains evolved significantly higher fitness than asexual controls in harsh environments (where beneficial mutations are thought to be more frequent), whereas fitness did not change for either strain in a permissive environment. However, the experiment could not distinguish whether the relative success of the sexual strain was caused by a reduction of negative (that is, synergistic) epistasis between deleterious mutations through segregation, or the recombination of beneficial mutations into a single strain.

In T4 bacteriophages the amount of epistasis between genes is significantly affected by the rate of recombination[58], hinting that recombination might unlink genes and promote modularity. Misevic, Ofria and Lenski[59] tested this hypothesis directly by measuring the degree and directionality of epistasis, as well as the modularity of genes, in both asexual and sexual populations of digital organisms. Digital populations that underwent recombination after replication evolved significantly faster, and were more robust to mutations. At the same time, although the fraction of pairs of substitutions that were epistatic did not differ between sexuals and asexuals, the fraction that were synergistic (aggravating rather than alleviating) was significantly higher in asexuals. They used a functional genomic array to study whether diminished aggravating epistasis is due to the evolution of modular genes. They found that modules (defined as code segments that are involved in a particular function) overlapped significantly less in sexual populations, which also had a larger distance between modules, on average. These findings indicate that recombination has a similar effect on the genome as high mutation rates: the average deleterious effect of a mutation, as well as the rate of synergistic epistasis are reduced by coding genes in a more modular, and therefore robust, fashion. Both findings reinforce Kondrashov's hypothesis[18,19].

Experiments such as this indicate that the form of epistasis is crucial when considering recombination and its effect on fitness. In a series of experiments to determine whether sexual digital organisms can avoid Muller's ratchet[60], they were found to have lower fitness at larger population sizes than their asexual counterparts (even though the sexuals could better survive population bottlenecks). Epistasis could explain this: previous studies[22] have found that deleterious mutations are predominantly antagonistic in digital organisms ($\beta < 1$). This leads to the accumulation of deleterious mutations, which asexuals can only get rid of 'one mutation per genome', whereas sexuals can remove several mutations per genetic death if they have been concentrated into one genome through recombination.

How could a transition from an asexual to sexual mode of reproduction then take place? We have seen that both high mutation rate and sexual recombination favour larger β values (deleterious mutations that interact less antagonistally), but high mutation rate alone cannot lead to $\beta > 1$, nor can the sexual mode be maintained if $\beta < 1$. It is tempting to think that genetic redundancy[61] would favour $\beta > 1$ because the first few mutations would lead to only small (or no) changes in fitness, whereas the fitness must drop catastrophically when the last redundant gene is knocked out. But this cannot be a path to sex, because redundancy can only be maintained in a sexual population. Indeed, if losing one copy of a redundant gene does not affect fitness, then redundancy cannot be maintained by asexuals. Therefore, sex seems to be a prerequisite of sex, a classic Catch-22 situation.

Perhaps this conclusion can be avoided with a pluralist approach that considers scenarios in which there

Fitness landscape
A visualization of the relationship between genotypes (providing the domain) and phenotype (the fitness) in evolutionary dynamics, for which the fitness can be characterized by a single real value, such as the rate of replication.

Modularity
In computer science, the degree to which a program is structured in independent components that can be moved around or modified without having an effect on other components (modules). In genetics, the degree to which a function is carried out by independent genes.

Twofold cost of sex
In a sexual population with two sexes, the twofold growth disadvantage of a population that has a 1:1 sex ratio, which is due to the fact that only females give birth. The opposite occurs when a population comprises only self-fertilizing females.

Muller's ratchet
In population genetics, the irreversible loss of alleles due to chance in small populations that reproduce asexually.

Red-Queen effect
The theoretical result of continuing evolutionary competition between host and parasite genes (or sometimes between competing mutations in clonal interference) that requires continued evolutionary innovation to survive.

are multiple mechanisms that favour sex. For example, parasites have been implicated in the evolution of sex through the Red-Queen effect[62,63]: an evolutionary arms race between competing and co-adapting strains[64–67]. Owing to the constantly changing environment, the recombination of beneficial mutations is believed to outweigh the load of combined deleterious mutations, so favouring sex. The Red-Queen effect also exists for digital organisms[68] (albeit as competition between beneficial mutants rather than between hosts and parasites). It is conceivable that the introduction of parasites will make multiple mutations comparatively more deleterious so that synergistic epistasis results, although it has not been possible so far to test this in digital organisms.

Conclusions

As highlighted by the studies described above, digital genetics has enriched classical evolutionary, functional and population genetics. In the future, modifications of digital organisms will allow their use to address questions that cannot currently be tackled. The evolution of gene regulation could be addressed — for example, by using variable transcription levels that are simulated by allowing multiple instruction pointers per gene that can run at different speeds. Also, work has begun to allow organisms to select a mate before recombination, allowing the study of another aspect of sexual reproduction. Furthermore, a satisfactory interaction of parasitic organisms with digital hosts is an important aim for the future: perhaps surprisingly, this is still problematic in Avida, as digital organisms seem to quickly evolve immunity to any parasite that is introduced, precluding the study of prolonged co-evolution (C. Ofria, personal communication). Finally, besides the areas mentioned in this review, it is worth noting that digital organisms have also been used in the study of processes such as the evolution of ecologies and the formation of species[69,70], indicating the potential of using this tool to understand diverse areas of evolutionary biology.

The first steps towards realizing the experimental evolutionist's dream were taken by Lenski, who initiated an evolution experiment using *Escherichia coli*[71–73] that has run uninterrupted for over 30,000 generations. Digital organisms allow even more fantastic dreams, because their exceptionally short generation times allow us to evolve complex genes — to quote Darwin[36] — "from so simple a beginning", and study their genetics. Digital organisms offer a glimpse of the basic patterns by which genetics and environment interact to produce these "endless forms most beautiful and most wonderful"[36].

1. Lenski, R. E. & Elena, S. F. Evolution experiments with microorganisms: The genetic bases of adaptation. *Nature Rev. Genet.* **4**, 457–469 (2003).
A thorough review of the state of the art of evolution experiments with microorganisms, with an emphasis on long-term experiments.
2. Wilke, C. O. & Adami, C. The biology of digital organisms. *Trends Ecol. Evol.* **17**, 528–532 (2002).
3. Zimmer, C. Testing Darwin. *Discover* **26**, 28–35 (2005).
4. Maynard Smith, J. Byte-sized evolution. *Nature* **335**, 772–772 (1992).
5. Dennett, D. in *Encyclopedia of Evolution* (ed. Pagel, M.) E83–E92 (Oxford Univ. Press, Oxford, 2002).
6. Ray, T. in *Proceedings of Artificial Life II* (eds Langton, C. G., Taylor, C., Farmer, J. D. & Rasmussen, S.) 371 (Addison Welsey, Redwood City, 1991).
This paper introduced the construction of digital life, and studied the interaction between host and parasite programmes, as well as the evolution of an arms race between them. Many of the design features of the Tierra system introduced here were adopted in the Avida system.
7. Adami, C. & Brown, C. T. in *Proceedings in Artificial Life IV* (eds Brooks, R. & Maes, P.) 377–381 (MIT Press, Boston, 1994).
8. Adami, C. *Introduction to Artificial Life* (Springer, New York, 1998).
9. Ofria, C. & Wilke, C. O. Avida: A software platform for research in computational evolutionary biology. *Artificial Life* **10**, 145–156 (2004).
This article is a user's guide for digital geneticists, outlining many of the options that are available to the experimenter along with a description of analysis tools.
10. Ofria, C., Adami, C. & Collier, T. C. Design of evolvable computer languages. *IEEE Trans. Evol. Comp.* **6**, 420–424 (2002).
11. Guo, H. H., Choe, J. & Loeb, L. A. Protein tolerance to random amino acid change. *Proc. Natl Acad. Sci. USA* **101**, 9205–9210 (2004).
12. Bloom, J. D. *et al.* Thermodynamic prediction of protein neutrality. *Proc. Natl Acad. Sci. USA* **102**, 606–611 (2005).
13. Atwood, K. C., Schneider, L. K. & Ryan, F. J. Periodic selection in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **37**, 146–155 (1951).
14. Yedid, G. & Bell, G. Microevolution in an electronic microcosm. *Am. Nat.* **157**, 465–487 (2001).
15. Gerrish, P. J. & Lenski, R. E. The fate of competing beneficial mutations in asexual populations. *Genetica* **102/103**, 127–144 (1998).
16. Wilke, C. O., Wang, J. L., Ofria, C., Lenski, R. E. & Adami, C. Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature* **412**, 331–333 (2001).
The first experiment that shows 'selection for robustness' as an active evolutionary pressure at high mutation rates. This analysis has spawned multiple attempts to find the 'survival of the flattest' effect in virus and viroid evolution experiments.
17. Wolf, J. B., Brodie, E. D. & Wade, M. J. *Epistasis and the Evolutionary Process* (Oxford Univ. Press, Oxford, 2000).
18. Kondrashov, A. S. Deleterious mutations and the evolution of sexual reproduction. *Nature* **366**, 435–440 (1988).
19. Kondrashov, A. S. Classification of hypotheses on the advantage of amphimixis. *J. Hered.* **84**, 372–387 (1993).
The ground-breaking hypothesis that now underlies most genetic theories of the origin and maintenance of the sexual mode of replication is first introduced here.
20. West, S. A., Peters, A. D. & Barton, N. H. Testing for epistasis between deleterious mutations. *Genetics* **149**, 435–444 (1998).
21. Rivero, A., Balloux, F. & West, S. A. Testing for epistasis between deleterious mutations in a parasitoid wasp. *Evolution* **57**, 1698–1703 (2003).
22. Lenski, R. E., Ofria, C., Collier, T. C. & Adami, C. Genome complexity, robustness and genetic interactions in digital organisms. *Nature* **400**, 661–664 (1999).
23. Wilke, C. O. & Adami, C. Interaction between directional epistasis and average mutational effects. *Proc. R. Soc. Lond. B* **268**, 1469–1474 (2001).
24. San Juan, R., Moya, A. & Elena, S. The contribution of epistasis to the architecture of fitness in an RNA virus. *Proc. Natl Acad. Sci. USA* **101**, 15376–15379 (2005).
25. Peck, J. R. A ruby in the rubbish: Beneficial mutations, deleterious mutations, and the evolution of sex. *Genetics* **137**, 597–606 (1994).
26. Ofria, C., Adami, C. & Collier, T. C. Selective pressures on genomes in molecular evolution. *J. Theor. Biol.* **222**, 477–483 (2003).
27. Kimura, M. & Maruyama, T. The mutational load with epistatic gene interactions in fitness. *Genetics* **54**, 1337–1351 (1966).
28. Eigen, M. Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**, 465–523 (1971).
29. Eigen, M. & Schuster, P. *The Hypercycle: A Principle of Natural Self-Organization* (Springer, Berlin, 1979).
30. Schuster, P. & Swetina, J. Stationary mutant distributions and evolutionary optimization. *Bull. Math. Biol.* **50**, 636–660 (1998).
31. Wilke, C. O. Probability of fixation of an advantageous mutant in a viral quasispecies. *Genetics* **162**, 467–474 (2003).
32. Wilke, C. O. Quasispecies theory in the context of population genetics. *BMC Evol. Biol.* **5**, 44 (2005).
33. van Nimwegen, E., Crutchfield, J. P. & Huynen, M. Neutral evolution of mutational robustness. *Proc. Natl Acad. Sci. USA* **96**, 9716–9720 (1999).
34. Wilke, C. O. & Adami, C. Evolution of mutational robustness. *Mutat. Res.* **522**, 3–11 (2003).
35. Dawkins, R. *Climbing Mount Improbable* (W. W. Norton, 1996).
A defence of the position that complex genes can evolve gradually, with an entire chapter devoted to the evolution of the eye.
36. Darwin, C. *On the Origin of Species* (John Murray, London, 1859).
37. Salvini-Plawen, L. V. & Mayr, E. On the evolution of photoreceptors and eyes. *Evol. Biol.* **10**, 207–263 (1977).
38. Land, M. F. & Fernald, R. D. The evolution of eyes. *Annu. Rev. Neurosci.* **15**, 1–29 (1992).
39. Yokoyama, S. & Radlwimmer, F. B. The molecular genetics and evolution of red and green colour vision in vertebrates. *Genetics* **158**, 1697–1710 (2001).
40. McShea, D. W. Metazoan complexity and evolution: Is there a trend? *Evolution* **50**, 477–492 (1996).
41. Adami, C. What is complexity? *Bioessays* **24**, 1085–1094 (2002).
42. Valentine, J. W., Collins, A. G. & Porter Meyer, C. Morphological complexitty increase in metazoans. *Paleobiology* **20**, 131–142 (1994).
43. McShea, D. W. Functional complexity in organisms: parts as proxies. *Biol. Philos.* **15**, 641–668 (2000).

44. Lenski, R. E., Ofria, C., Pennock, R. T. & Adami, C. The evolutionary origin of complex features. *Nature* **423**, 139–144 (2003).
The first detailed analysis of an evolutionary path that leads to a complex gene, following the emergence of the EQU gene in 50 replicate populations of digital organisms, mutation by mutation.
45. Johnson, T. & Barton, N. H. The effect of deleterious alleles on adaptation in asexual organisms. *Genetics* **162**, 395–411 (2002).
46. Wagner, G. P. & Altenberg, L. Complex adaptations and the evolution of evolvability. *Evolution* **50**, 967–976 (1996).
47. Kirschner, M. & Gerhart, J. Evolvability. *Proc. Natl Acad. Sci. USA* **95**, 8420–8427 (1998).
48. Wagner, A. *Robustness and Evolvability in Living Systems* (Princeton Univ. Press, Princeton, 2005).
49. Caporale, L. Chance favors the prepared genome. *Ann. NY Acad. Sci.* **870**, 1–21 (1999).
50. Edlund, J. A. & Adami, C. Evolution of robustness in digital organisms. *Artificial Life* **10**, 167–79 (2004).
51. Barrell, B. G., Air, G. M. & Hutchison, C. A. Overlapping genes in bacteriophage ϕX174. *Nature* **264**, 34–41 (1976).
52. Miyata, T. & Yasunaga, T. Evolution of overlapping genes. *Nature* **272**, 532–535 (1978).
53. Johnson, Z. I. & Chisholm, S. W. Properties of overlapping genes are conserved in microbial genomes. *Genome Res.* **14**, 2268–2272 (2004).
54. Krakauer, D. C. Stability and evolution of overlapping genes. *Evolution* **54**, 731–739 (2000).
55. Ofria, C. & Adami, C. in *Evolution as Computation*, *DIMACS Workshop* (eds Landweber, L. & Winfree, E.) 296–313 (Springer, New York, 2002).
56. Mizokami, M., Orito, E., Ohba, K., Lau, J. Y. N. & Gojobori, T. Constrained evolution with respect to gene overlap of hepatitis B virus. *J. Mol. Evol.* **44** (Suppl. 1), S83–S90 (1997).
57. Goddard, M. R., Godfray, H. C. J. & Burt, A. Sex increases the efficacy of natural selection in experimental yeast populations. *Nature* **434**, 636–640 (2005).
58. Malmberg, R. L. The evolution of epistasis and the advantage of recombination in populations of bacteriophage T4. *Genetics* **86**, 607–621 (1977).
59. Misevic, D., Ofria, C. & Lenski, R. E. Sexual reproduction reshapes the genetic architecture of digital organisms. *Proc. R. Soc. B.* 8 November 2005 (doi:10.1098/rspb.2005.3338).
An ingenious study of the emergence of modules as a response to code recombination in digital organisms.
60. Misevic, D., Lenski, R. E. & Ofria, C. in *Proceedings in Artificial Life IX* (eds Pollack, J., Bedau, M. A., Husbands, P., Ikegami, T. & Watson, R.) 340–345 (MIT Press, Boston, 2004).
61. Krakauer, D. C. & Plotkin, J. B. Redundancy, antiredundancy, and the robustness of genomes. *Proc. Natl Acad. Sci. USA* **99**, 1405–1409 (2002).
62. van Valen, L. A new evolutionary law. *Evol. Theory* **1**, 1–30 (1973).
63. Ridley, M. *The Red Queen: Sex and the Evolution of Human Nature* (Macmillan, New York, 1993).
64. Maynard Smith, J. *The Evolution of Sex* (Cambridge Univ. Press, 1978).
65. Bell, G. *The Masterpiece of Nature: The Evolution and Genetics of Sexuality* (Univ. California Press, Berkeley, 1982).
66. Lively, C. M. & Howard, R. S. Selection by parasites for clonal diversity and mixed mating. *Phil. Trans. R. Soc. Lond. B* **346**, 271–281 (1994).
67. Howard, R. S. & Lively, C. M. The maintenance of sex by parasitism and mutation accumulation under epistatic fitness functions. *Evolution* **52**, 604–610 (1998).
68. Wilke, C. O. Does the Red Queen reign in the kingdom of digital organisms? *Lect. Notes Artificial Intell.* **2801**, 405–414 (2003).
69. Cooper T. & Ofria, C. in *Proceedings in Artificial Life VIII* (eds Standish, R. K., Bedau, M. A. & Abbass, H. A.) 227–232 (MIT Press, Cambridge, 2002).
70. Chow, S. S., Wilke, C. O., Lenski, R. E., Ofria, C. & Adami, C. Adaptive radiation from resource competition in digital organisms. *Science* **305**, 83–85 (2004).
A study of the selective pressures that give rise to stably co-existing 'ecotypes', by varying resource availability to force negative frequency-dependent selection.
71. Lenski, R. E. & Travisano, M. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proc. Natl Acad. Sci. USA* **91**, 6808–6814 (1994).
This landmark paper introduces the modern experimental evolution philosophy of replicate populations evolving for thousands of generations in controlled environments, with accurate fitness measurements and a perfect 'fossil record'.
72. Lenski, R. E., Rose, M. R., Simpson, S. C. & Tadler, S. C. Long-term experimental evolution in *Escherichia coli*: adaptation and divergence during 2,000 generations. *Am. Nat.* **138**, 1315–1341 (1991).
73. Lenski, R. E. Phenotypic and genomic evolution during a 20,000-generation experiment with the bacterium *Escherichia coli*. *Plant Breed. Rev.* **24**, 225–265 (2004).
74. Gould, S. J. *Wonderful Life: The Burgess Shale and the Nature of History* (W. W. Norton, New York, 1989).
75. Travisano, M., Mongold, J. A., Bennett, A. F. & Lenski, R. E. Experimental tests of the roles of adaptation, chance, and history in evolution. *Science* **267**, 87–90 (1995).
Another landmark study that introduces statistical tests for disentangling the effects of history, chance and adaptation on evolution using microorganisms.
76. Wagenaar, D. A. & Adami, C. Influence of chance, history, and adaptation on digital evolution. *Artificial Life* **10**, 181–190 (2004).
77. Fontana, W. & Buss, L. W. What would be conserved 'if the tape were played twice'? *Proc. Natl Acad. Sci. USA* **91**, 757–761 (1994).
78. McLean, R. C. & Bell, G. Divergent evolution during an experimental adaptive radiation. *Proc. R. Soc. Lond. B* **270**, 1645–1650 (2003).
79. Coyne, J. & Orr, H. A. *Speciation* (Sinauer, Sunderland, 2004).
80. Rundle, H. D. & Nosil, P. Ecological speciation. *Ecol. Lett.* **8**, 336–352 (2005).
81. Otto, S. P. & Lenormand, T. Resolving the paradox of sex and recombination. *Nature Rev. Genet.* **3**, 252–261 (2002).
82. Rice, W. R. Experimental tests of the adaptive significance of sexual recombination. *Nature Rev. Genet.* **3**, 241–251 (2002).
83. Michod, R. E. & Levin, B. R. (eds) *The Evolution of Sex: A Critical Review of Current Ideas* (Sinauer, Sunderland, 1988).
84. Gould, S. J. *Full House: The Spread of Excellence from Plato to Darwin* (Harmony Books, New York, 1996).
85. Bonner, J. T. *The Evolution of Complexity* (Princeton Univ. Press, Princeton, 1988).
86. Dyson, F. *Origins of Life* (Cambridge Univ. Press, Cambridge, 1999).
87. Zimmer, C. How and where did life on Earth arise? *Science* **309**, 89 (2005).
88. Rasmussen, S., Chen, L., Stadler, B. & Stadler, P. F. Proto-organism kinetics: evolutionary dynamics of lipid aggregates with genes and metabolism. *Orig. Life Evol. Biosph.* **34**, 171–180 (2004).
89. Segré, D., Ben-Eli, D. & Lancet, D. Compositional genomes: Prebiotic information transfer in mutually catalytic noncovalent assemblies. *Proc. Natl Acad. Sci. USA* **97**, 4112–4117 (2000).
90. Rasmussen, S. *et al.* Transitions from nonliving to living matter. *Science* **303**, 963–965 (2004).
91. Maynard Smith, J. & Szathmary, E. *The Major Transitions in Evolution* (Oxford Univ. Press, Oxford, 1995).
92. Drake, J. W. & Holland, J. J. Mutation rates among RNA viruses. *Proc. Natl Acad. Sci. USA* **96**, 13910–13913 (1999).
A thorough compilation of mutation rates in RNA viruses that indicates that evolution optimizes the product of per-site mutation rate and genome length, but not each independently.
93. Gause, G. F. *The Struggle for Existence* (Williams & Wilkins, Baltimore, 1934).

Acknowledgements
I would like to thank R.E. Lenski, D. Misevic, C. Ofria, and C.O. Wilke for extensive discussions. I am further indebted to D. Misevic for kindly providing me with figure 1b, and D. Weise for the illustration in box 3.

Competing interests statement
The author declares no competing financial interests.

FURTHER INFORMATION
Avida Digital Life Platform: http://sourceforge.net/projects/avida
Devolab — the digital evolution laboratory at Michigan State University: http://devolab.cse.msu.edu
Index of Avida Documentation: http://devolab.cse.msu.edu/software/avida/doc
The Digital Life Laboratory, California Institute of Technology: http://www.dllab.caltech.edu
Access to this interactive links box is free online.