What evolutionary biologists can learn from Artificial Life

Santiago F. Elena
Evolutionary Systems Virology Group

Big questions in Evolutionary Biology and experimental limitations

- ✓ The evolution of complex traits.
- ✓ The role of neutral variation in adaptive evolution.
- ✓ Selection for fitness *vs* selection for robustness.
- ✓ The topography of adaptive landscapes and the evolution of landscapes.
- ✓ Eco-evolutionary dynamics: how evolution changes ecology and how ecology modulates evolution.
- ✓ Evolution of phenotype-genotype maps.
- ✓ The evolution of genetic systems (sex, speciation, genome architecture).

The advantages of microbial Experimental Evolution

✓ They are easy to propagate and enumerate.

✓ They reproduce quickly, which allows experiments to run for many generations.

✓ They allow large populations in small spaces, which facilitates experimental replication.

✓ They can be stored in suspended animation and later revived, which allows the direct comparison of ancestral and evolved types.

✓ Many microbes reproduce asexually and the resulting clonality enhances the precision of experimental replication.

✓ Asexuality also maintains linkage between a genetic marker and the genomic background into which it is placed, which facilitates fitness measurements.

✓ It is easy to manipulate environmental variables, such as resources, as well as the genetic composition of founding populations.

✓ There are abundant molecular and genomic data for many species, as well as techniques for their precise genetic analysis and manipulation.

Peculiarities of RNA viruses

✓ High genetic variability. Orders of magnitude greater than for DNA-based organisms.

✓ High mutation rates: 2.5×10^{-4} s/s/r for VSV, 5×10^{-5} for TEV and 2.5×10^{-3} for CChMVd. Such mutation rates are consequence of the lack of proofreading mechanisms in viral RdRp.

✓ Compacted genome: 11162 nts for VSV and 9494 for TEV.

✓ Huge numbers of generations per time unit: ~10^3 PFU/cell in 6 - 8 hpi for VSV or ~10^6 LFU/g 5 dpi for TEV.

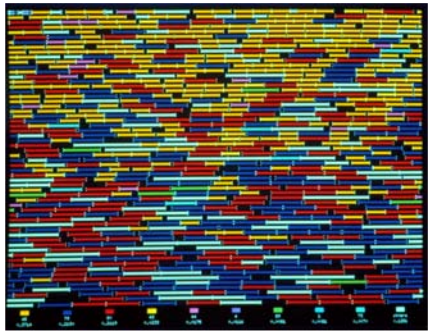
✓ The variability is a key factor for pathogenicity.

✓ It is impossible talking about a single defined entity. Instead we shall talk on a distribution of genomes centered around a more frequent one: *Quasispecies*.

✓ Relatively easy to map genotypes into phenotypic space.

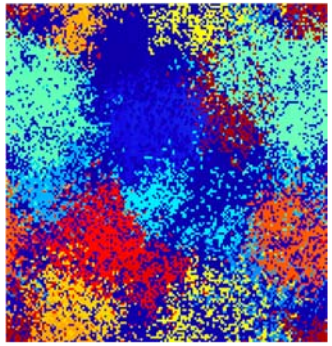
✓ Viral infectious diseases represent the most important threat to animal and plant health.

One step beyond Experimental Evolution: Artificial Life

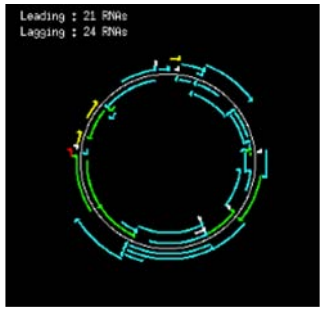



Thomas S. Ray, the **Tierra** project



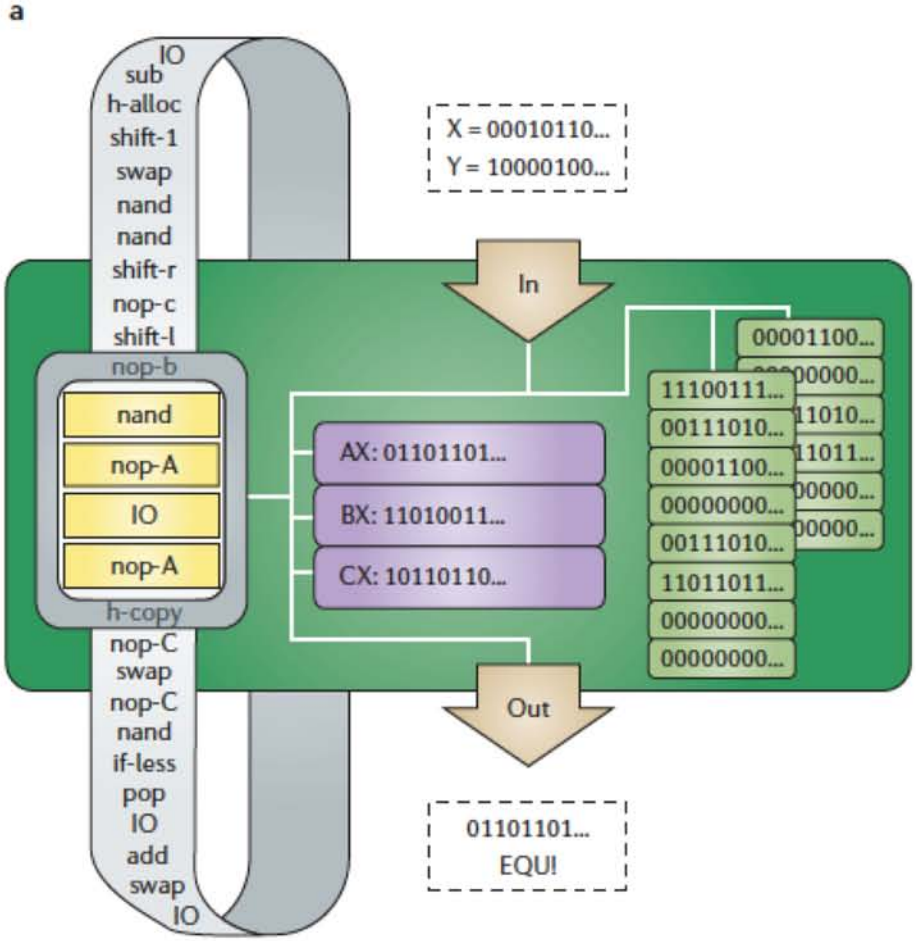


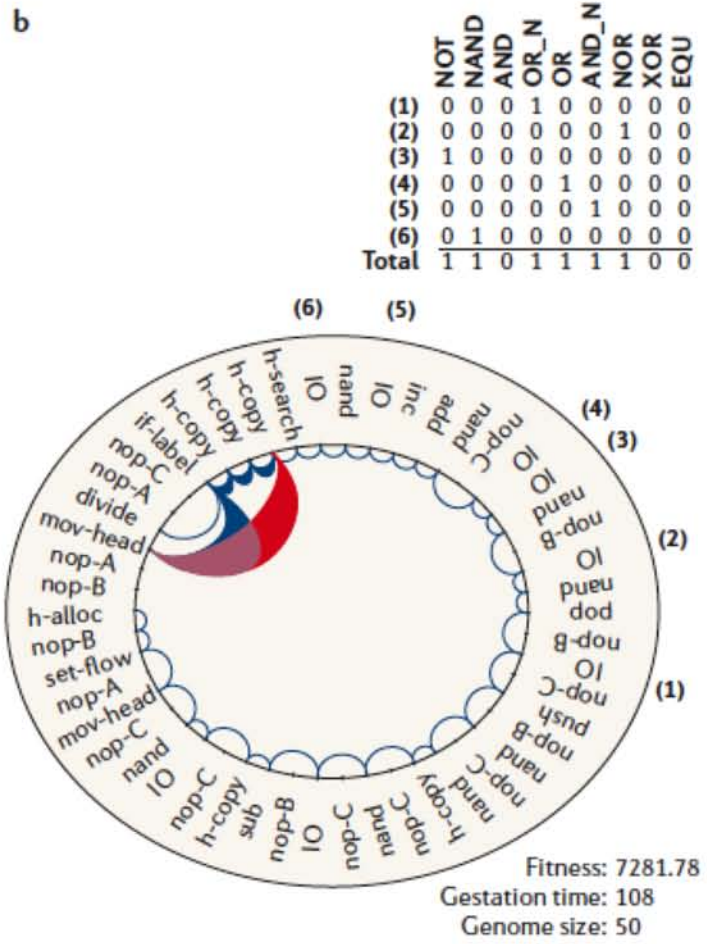
Chris Adami and Charles Ofria, the **Avida** platform

Guillaume Beslon, the **Aevol** platform

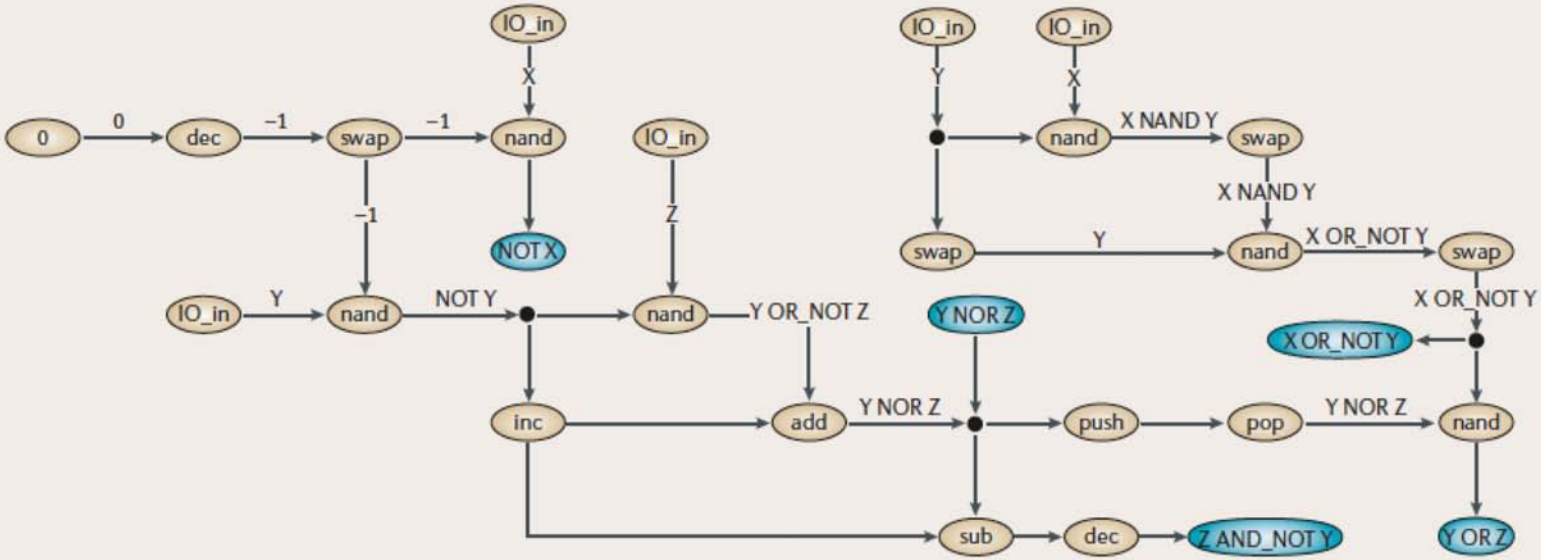
Avidians genomes





Avidians logical metabolism

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.

Adami C. 2006 *Nat. Rev. Genet.* **7**, 109-18

Functional genomics in Avida

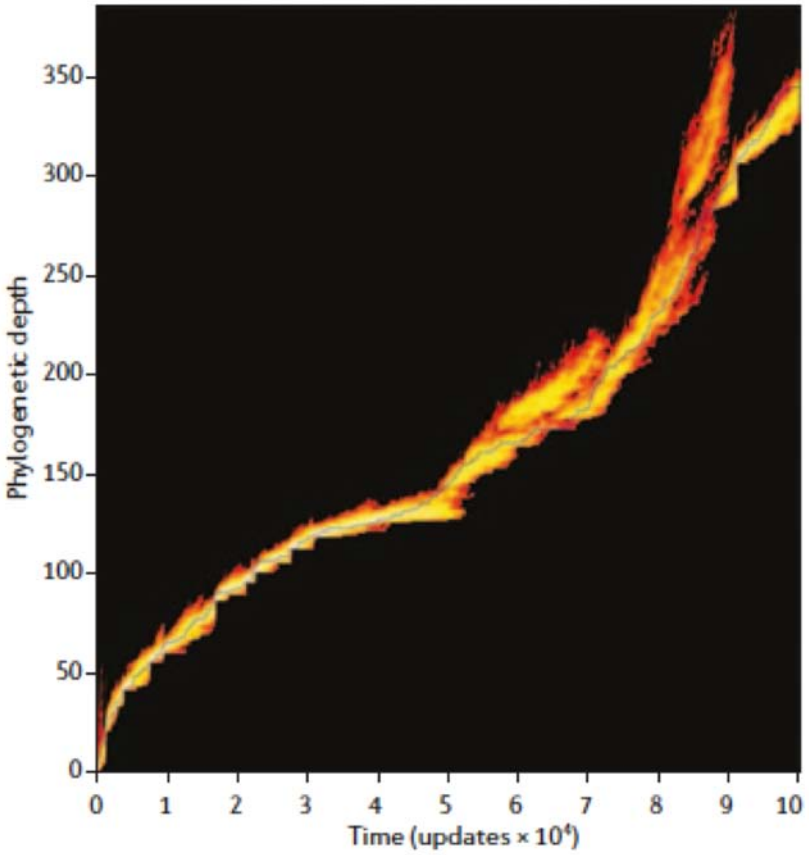
Box 4 | The functional genomic array of an avidian

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	15

Adami C. 2006 *Nat. Rev. Genet.* **7**, 109-18

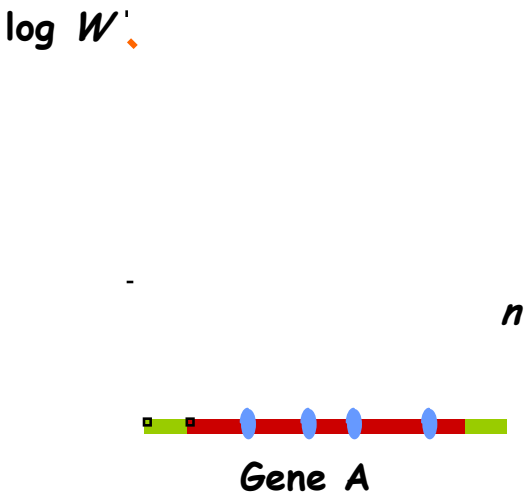
Phylogenetics in Avida



Adami C. 2006 *Nat. Rev. Genet.* **7**, 109-18

The evolution of genome complexity

Genome complexity and epistasis

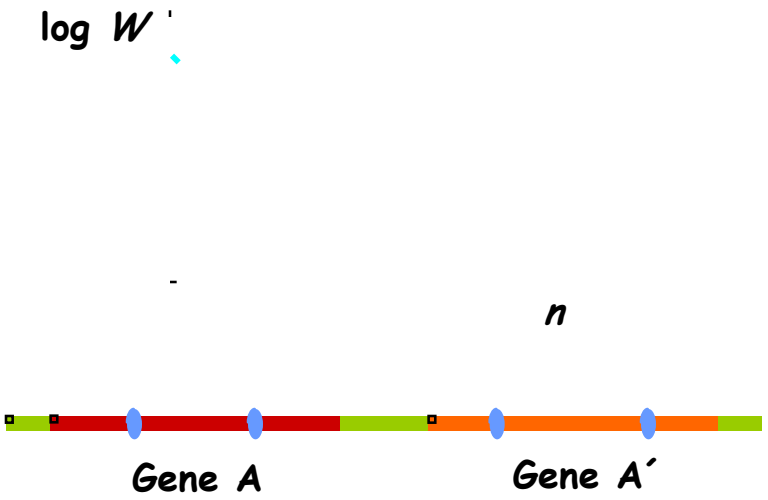


Anti-redundancy

Strong deleterious fitness effects

Antagonistic epistasis

Expected for RNA viruses



Redundancy

Mild deleterious fitness effects

Synergistic epistasis

Expected for complex organisms

Krakauer DK, Plotkin JB. 2002 *Proc. Natl. Acad. Sci. USA* **99**, 1405-9
Sanjuán R, Elena SF. 2006 *Proc. Natl. Acad. Sci. USA* **103**, 14402-5

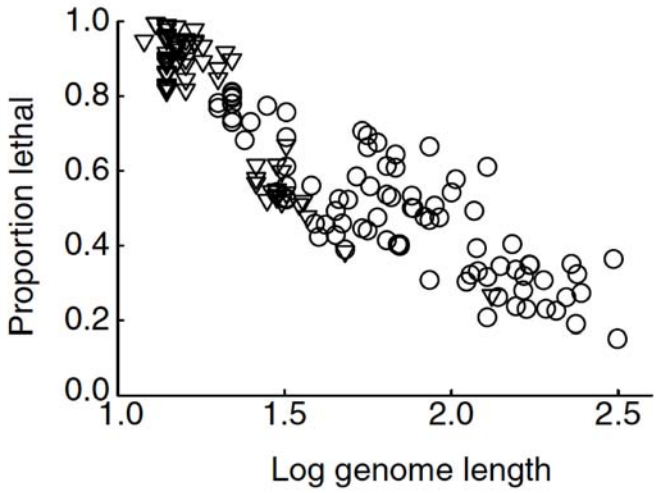


Figure 1 Proportion of single point mutations that are lethal for digital organisms. Shown as a function of $\log_{10}$-transformed genome length. Circles, complex organisms; triangles, simple organisms.

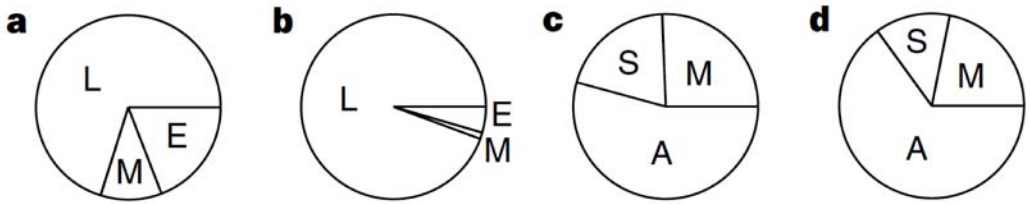


Figure 3 Proportions of mutational pairs classified according to their interaction. L, lethal: at least one mutation is lethal alone, as is the double mutant. M, multiplicative: neither mutation is lethal, and the relative fitness of the double mutant is exactly equal to the product of the relative fitnesses of the two single mutations. E, epistatic: the double mutant's fitness is unequal to the multiplicative expectation. S, synergistic: double mutant less fit than expected. A, antagonistic: the double mutant is more fit than expected. Average distributions are for complex organisms (**a**), simple organisms (**b**), complex excluding lethals (**c**), and simple excluding lethals (**d**).

$$\ln W(x) = -\alpha x^{\beta}$$

Table 1 Comparisons between complex and simple digital organisms of genome size and several mutational-effect parameters

Response variable	Mean complex (± s.d.)	Mean simple (± s.d.)	Mean difference (± s.d.)	*P**
Genome length	91.25 (69.07)	19.80 (14.18)	71.45 (64.76)	<0.0001
Decay test, α	0.581 (0.207)	1.141 (0.591)	–0.560 (0.562)	<0.0001
Decay test, β	0.896 (0.081)	0.972 (0.192)	–0.077 (0.201)	0.0011
Pair test, proportion epistatic of total	0.191 (0.093)	0.045 (0.080)	0.146 (0.122)	<0.0001
Pair test, proportion epistatic of non-lethal	0.743 (0.243)	0.781 (0.234)	–0.038 (0.303)	0.4374
Pair test, proportion synergistic of epistatic	0.271 (0.093)	0.168 (0.159)	0.103 (0.175)	<0.0001

* Two-tailed Wilcoxon signed-ranks test of the differences between 87 paired complex and simple organisms.

Fragile

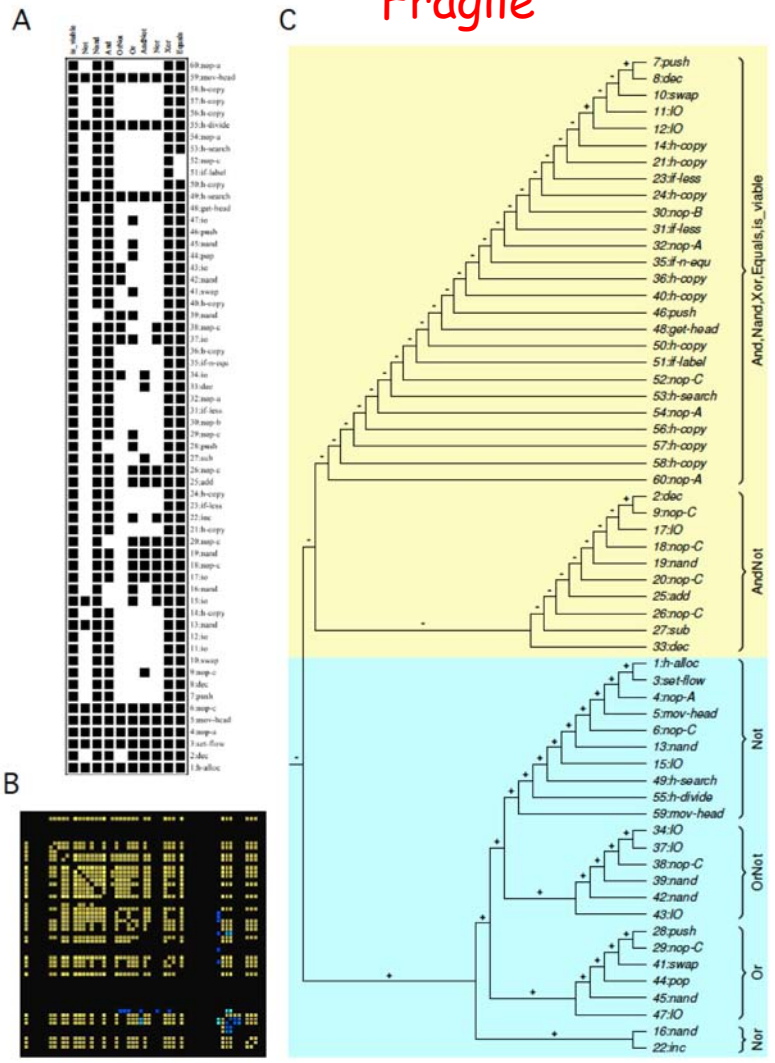


Figure 1: Analysis of functional modularity for a brittle avidian ($\langle \epsilon_\omega \rangle < 0$) with low modularity ($Q = 0.15$). (A) Task map showing the implication of each genomic instruction on the nine different tasks. (B) Heat-map illustrating the intensity of epistatic interactions between pairs of instructions in the genome. The stronger the blue, the more antagonistic (positive) epistasis; the stronger the yellow, the more synergistic (negative) epistasis. (C) Cladogram constructed from the epistasis matrix. Branches have been decorated with the average epistasis of the corresponding subtree (see text).

Robust

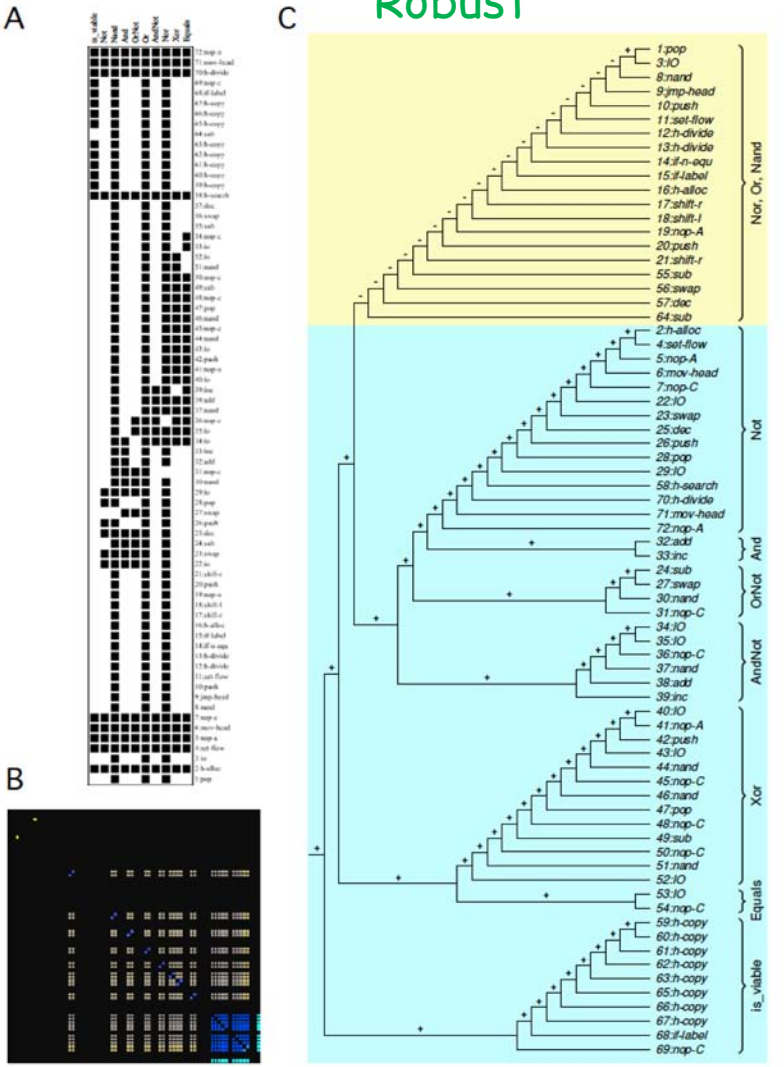
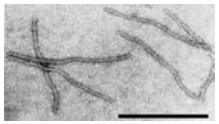


Figure 2: Analysis of functional modularity for a robust avidian ($\langle \epsilon_\omega \rangle > 0$) with high modularity ($Q = 0.21$). (A) Task map showing the implication of each genomic instruction on the nine different tasks. (B) Heat-map illustrating the intensity of epistatic interactions between pairs of instructions in the genome. The stronger the blue, the more antagonistic (positive) epistasis; the stronger the yellow, the more synergistic (negative) epistasis. (C) Cladogram constructed from the epistasis matrix. Branches have been decorated with the average epistasis of the corresponding subtree (see text).

Fragile organisms are characterized by **antagonistic non-modular epistasis**, while **robust** organisms are characterized by **synergistic negative epistasis**

Valverde S *et al.* 2012 *Artif. Life* **113**, 111-5

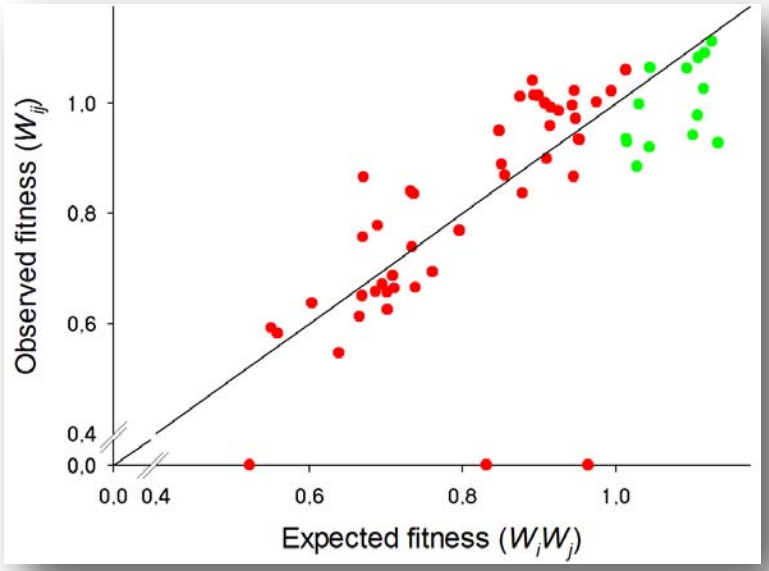
Therefore, we expect viruses to be fragile…



	Proportion	E(*s*)	Proportion	E(*s*)
Lethal	39.6%	–1	40.9%	–1
Deleterious	29.2%	–0.244	36.4%	–0.490
Neutral	27.1%	0	22.7%	0
Beneficial	4.2%	0.042	0.0%	-
Total	100% (48)	–0.476	100% (66)	–0.491

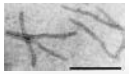
Sanjuán R *et al.* 2004 *Proc. Natl. Acad. Sci. USA* **101**, 8396-401
Carrasco P *et al.* 2007 *J. Virol.* **81**, 12979-84

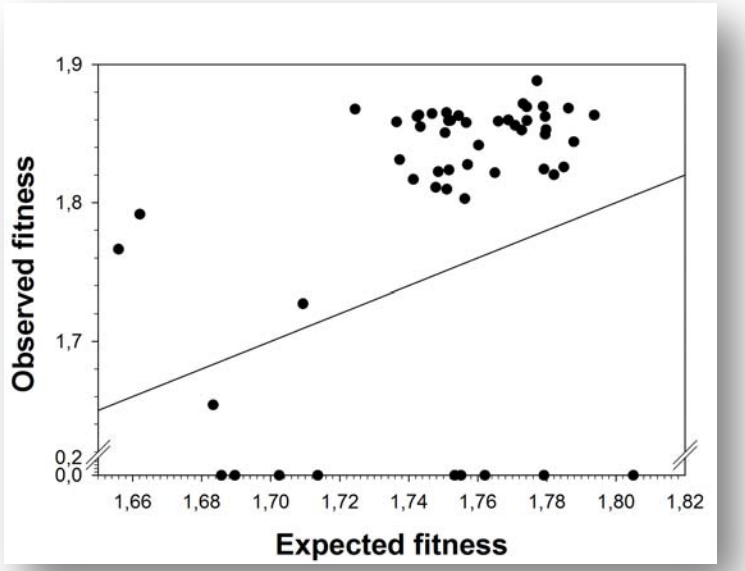
…and dominated by positive epistasis



- ✓ Among deleterious pairs: 3 cases of antagonistic epistasis
- ✓ 3 synthetic lethals (synergistic epistasis)
- ✓ $\langle\varepsilon\rangle = 0.034 \pm 0.010$ (t-test, $P = 0.002$)

- ✓ Among beneficial pairs: 6 cases of antagonistic epistasis,
- ✓ Including 5 cases of decompensatory epistasis.
- ✓ $\langle\varepsilon\rangle = -0.077 \pm 0.017$ (t-test, $P < 0.001$)



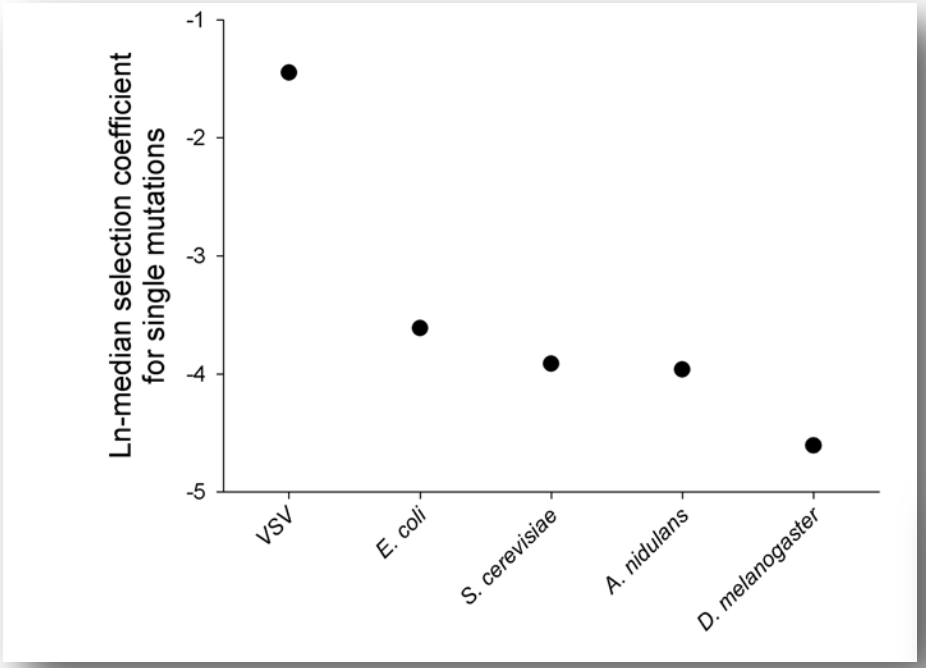


- ✓ 9 cases of synthetic lethals (synergistic epistasis).
- ✓ 11 cases of antagonistic epistasis.

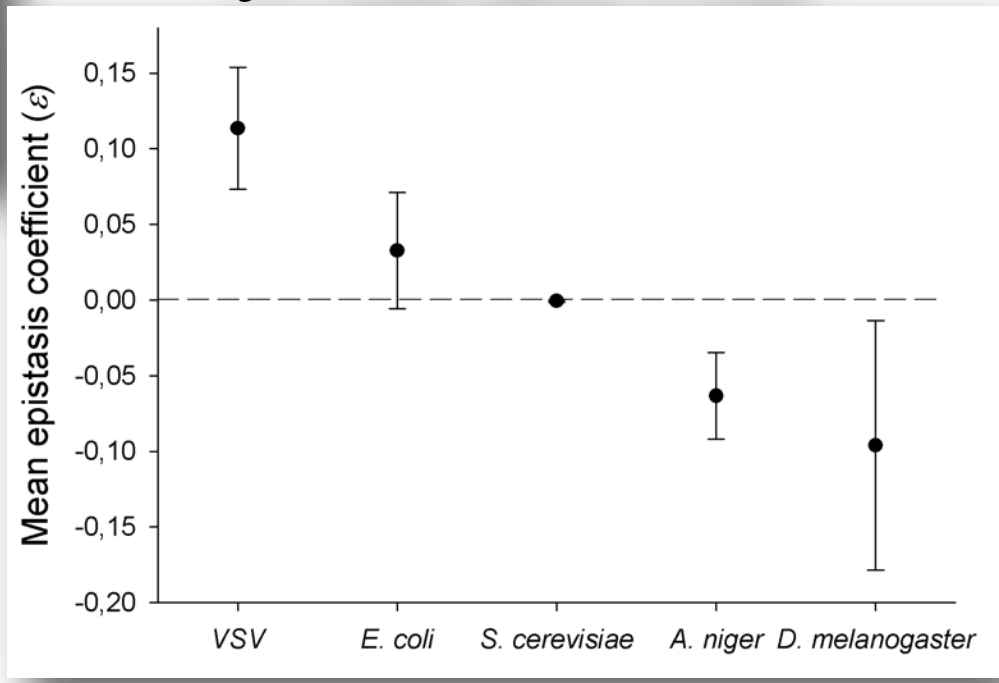
Sanjuán R *et al.* 2004 *Proc. Natl. Acad. Sci. USA* **101**, 15376-9
Lalić J, Elena SF. 2012 *Heredity* **109**, 71-7

In contrast to more complex organisms

Spearman's r_S = –1,000, 4 df, P < 0.001



r_S = –0.315, 249 df, P < 0.001



Sanjuán R, Elena SF. 2006 *Proc. Natl. Acad. Sci. USA* **103**, 14402-5

The origin of complex features

The obsession of intelligent designers: the evolution of a complex eye

- ✓ Darwinian explanation: complex features do not evolve *de novo* but via incremental transitions of functional states.

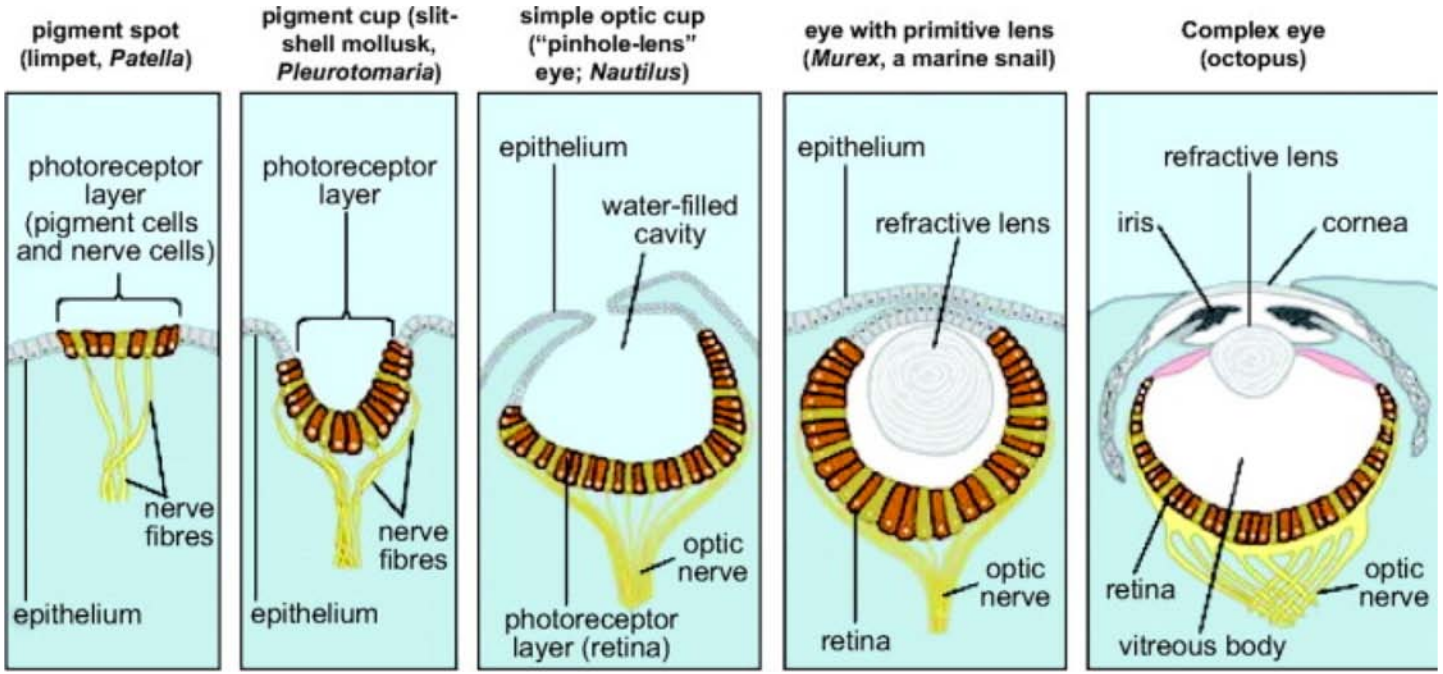


Table 1 **Rewards for performing nine one- and two-input logic functions**

Function name	Logic operation	Computational merit
NOT	~A; ~B	2
NAND	~(A and B)	2
AND	A and B	4
OR_N	(A or ~B); (~A or B)	4
OR	A or B	8
AND_N	(A and ~B); (~A and B)	8
NOR	~A and ~B	16
XOR	(A and ~B) or (~A and B)	16
EQU	(A and B) or (~A and ~B)	32

The symbol '~' denotes negation. The reward for computational merit increases with 2^n, where n is the minimum number of nand operations needed to perform the listed function. Symmetrical operations, shown separated by a semi-colon, are treated as the same function. No added benefit is obtained for performing any function multiple times. These functions include all one- and two-input logic operations except ECHO, which requires no nand operations and was not rewarded.

- ✓ Evolution started with an avidian only able of replicate but not performing any logical operation (L = 50).

✓ The resulting dominant genotype was 344 "genotypes" away from the ancestral (L = 83).

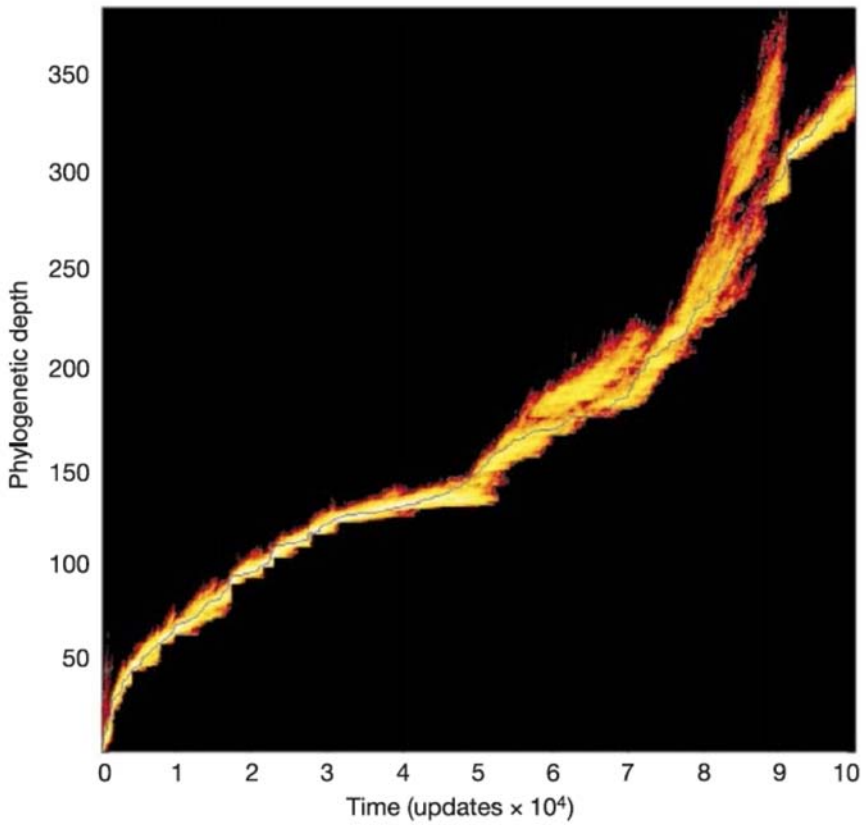


Figure 2 Phylogenetic depth versus time in the case-study population. Phylogenetic depth is the cumulative number of generations in which an organism's genotype differs from its parent. The exact line of descent leading to the most abundant final genotype is shown as the blue line. Colours indicate the relative abundance of genotypes at any depth, yellow being more abundant than red.

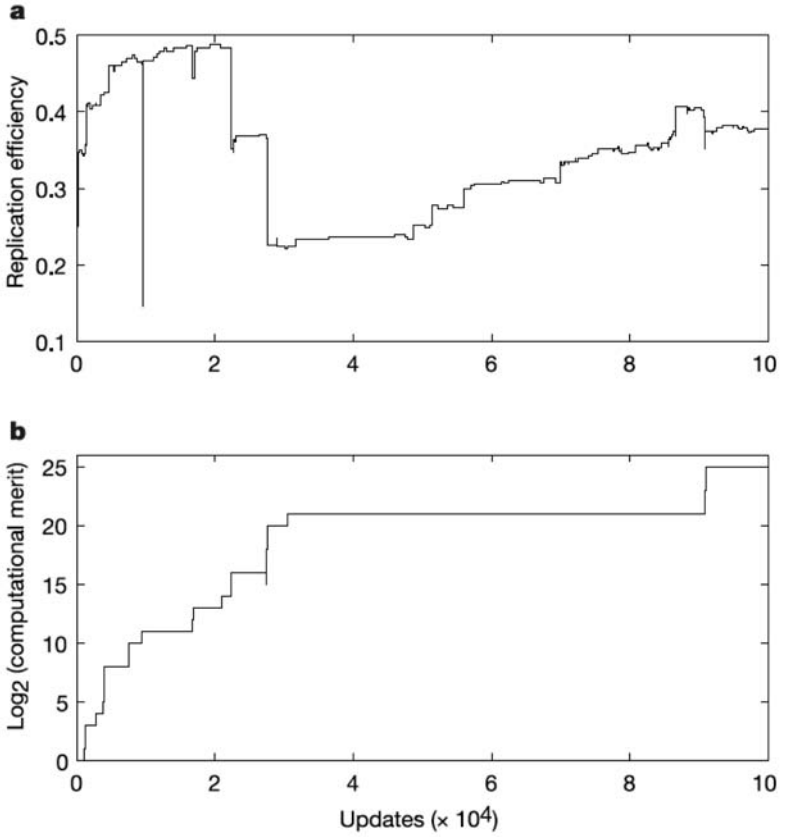


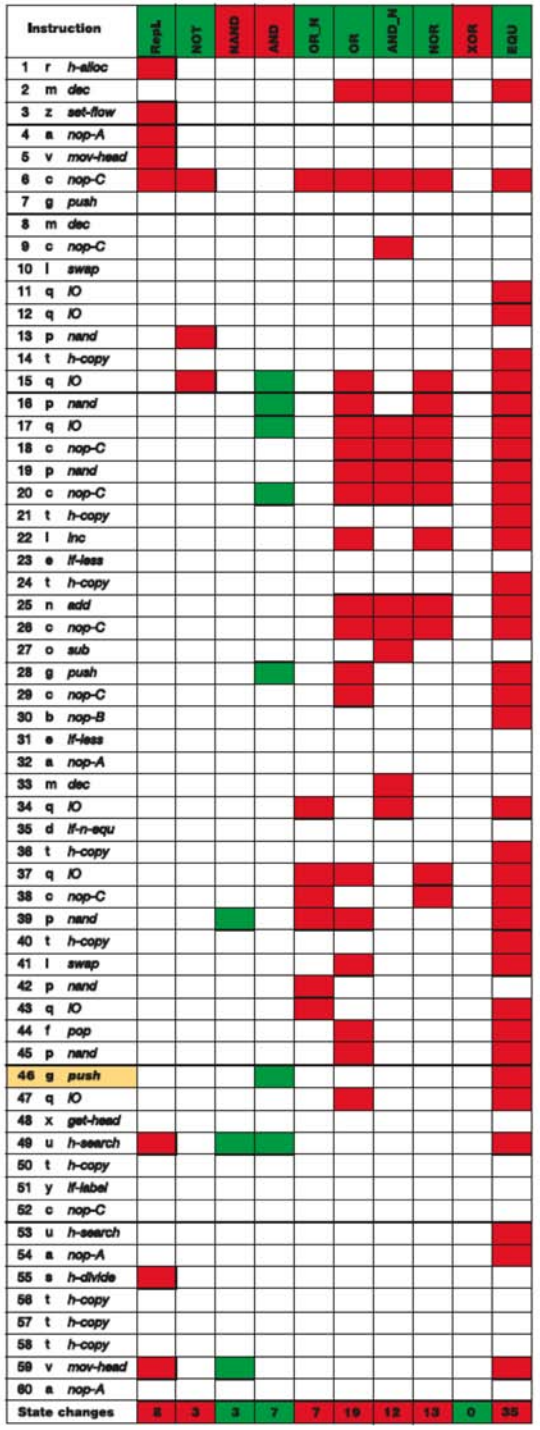
Figure 3 Trajectories for two fitness components, showing each genotype in the line of descent for the case-study population. **a**, Replication efficiency, which is the ratio of genome length to energy used in an organism's lifetime. **b**, Computational merit, shown $\log_2$-transformed, which is the product of all the rewards obtained by an organism for logic functions performed during its lifetime.

✓ EQU first appeared after 111 "genotypes": 103 single, 6 double and 2 triple mutations in the lineage.

✓ Some mutations were deleterious when first appeared, but highly beneficial afterwards.

The genome of the first avidian that did EQU

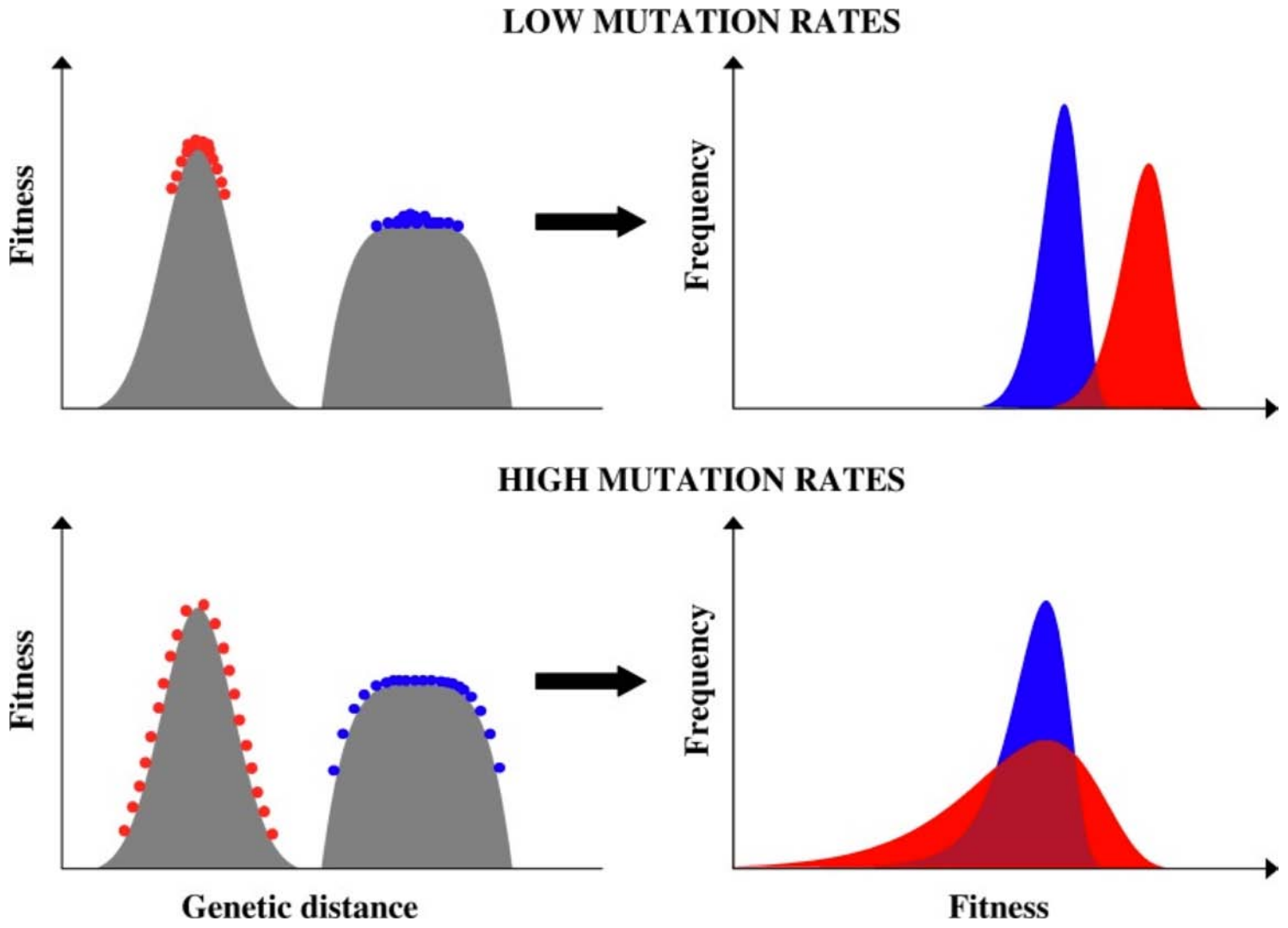
Figure 4 Functional-genomic array for the first organism to perform EQU in the case-study population. Its genome sequence is shown to the left; the instruction highlighted in yellow is the pivotal mutation that yielded EQU but simultaneously eliminated AND. Top labels denote replication (Repl.) and logic functions; associated colours show whether this organism can (green) or cannot (red) perform the function. The fill in each interior cell shows the effect on the function of replacing the instruction with a null instruction. Red, null mutation destroys existing function; blank, null mutation has no qualitative effect; green, null mutation produces new function. The number of state changes for each function is shown at the bottom.



Lenski RE *et al.* 2003 *Nature* **423**, 139-44

Mutation, neutrality, robustness, and the rise of the flattest

The quasispecies effect or "the survival of the flattest"



The survival of the flattest in Avida

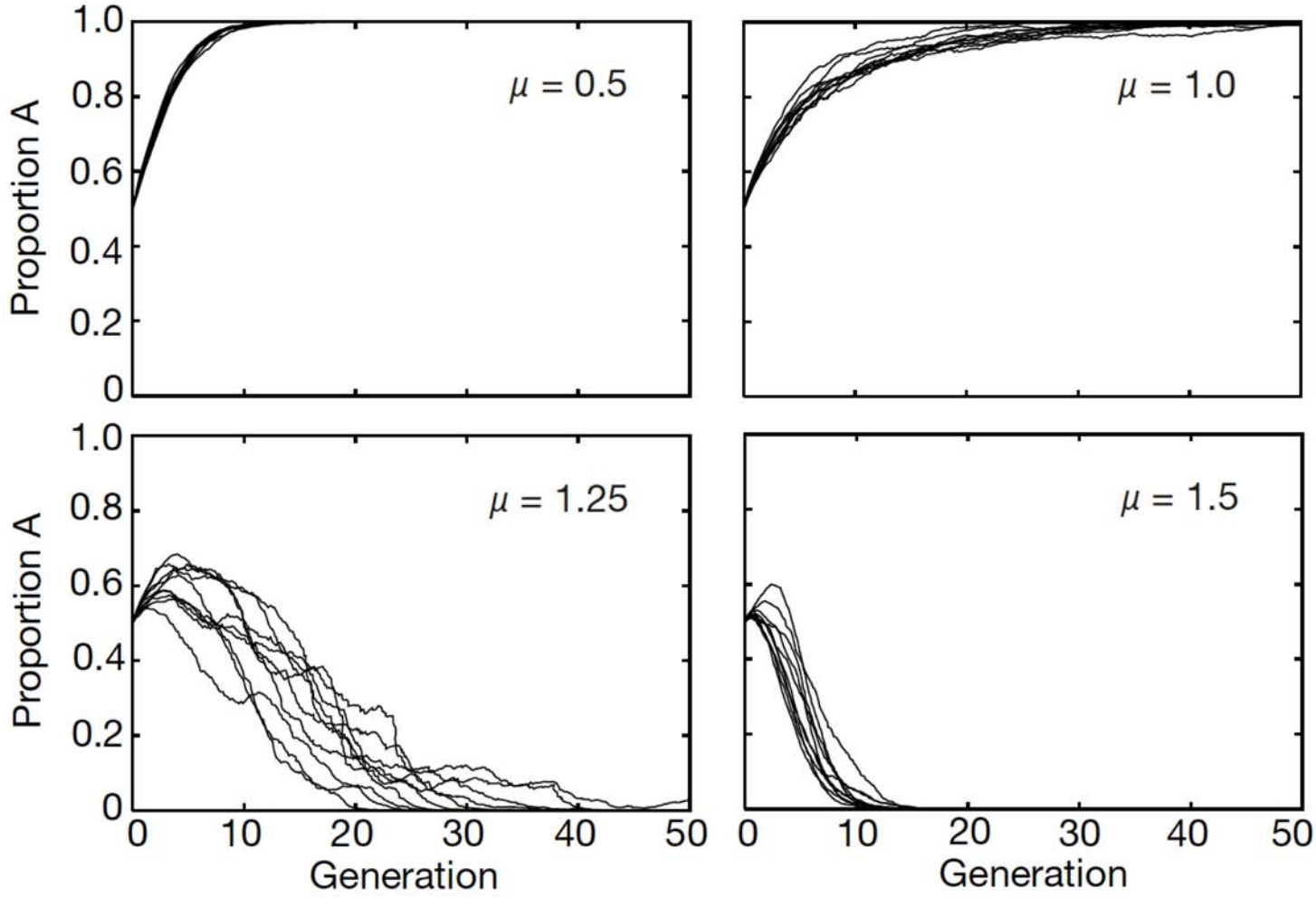


Figure 1 Competitions for one pair of organisms at four different mutation rates. Organism A replicates 1.96 times faster than B. μ, Genomic mutation rate.

Wilke *CO et al.* 2001 *Nature* **412**, 331-3

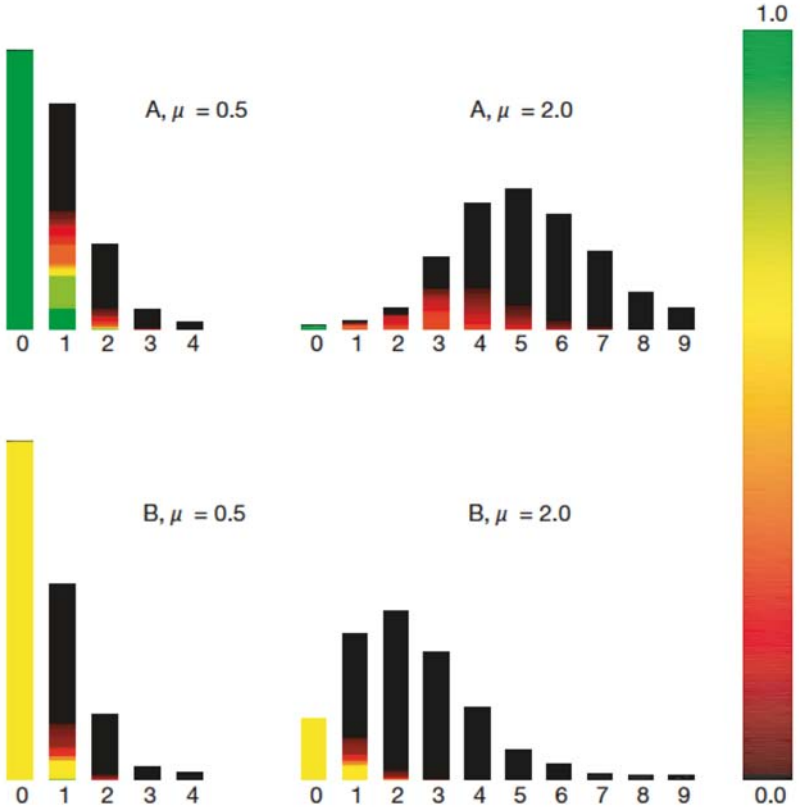


Figure 2 Genotype distributions after 15 generations from populations seeded uniformly with either A or B. Numbers below the bars are Hamming distances from the seeding genotype; bar heights are relative frequencies of genotypes at those distances. Colour coding indicates the intrinsic replication rates of all the genotypes. The colour scale was normalized to the intrinsic replication rate of A. The pair of organisms in this graph is the same pair for which the dynamics of competition were shown in Fig. 1. The right tails of the distributions are truncated for the purposes of illustration only.

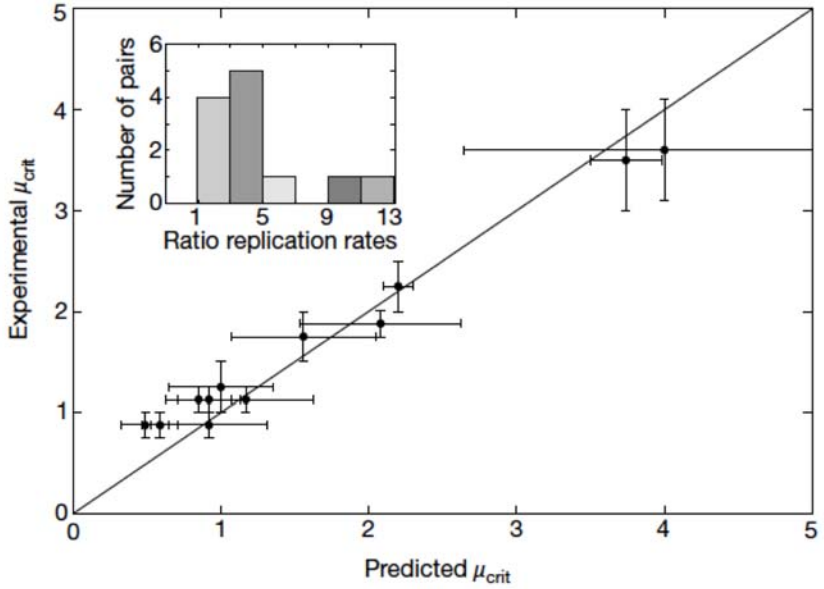


Figure 3 Critical mutation rates obtained from competition experiments versus predicted estimates based on each organism's intrinsic replication rate, w_0, and mutational robustness parameters, *a* and *b*. Points along the diagonal line would imply perfect agreement between experiment and prediction; the actual correlation coefficient is 0.991 ($P < 0.0001$). The inset shows the distribution of intrinsic replication rate ratios ($w_{0,A}/w_{0,B}$) for the 12 pairs.

Rates of evolution depend on the mutational supply μN

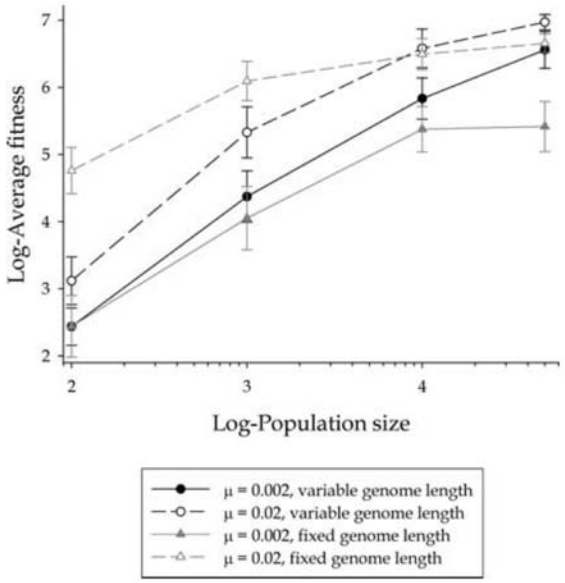


Figure 1. Effects of population size, mutation rate, and genome-length restriction on log-transformed fitness of the most abundant genotypes isolated at the end of the evolution experiments. Error bars represent standard errors of the mean ($n = 10$).

Table 1. Analysis of covariance of the log-transformed fitness data shown in Figure 1. Mutation rate (μ) and genome-length restriction (GLR) were treated as fixed factors; log-population size (log N) was treated as a covariate. Type III sums of squares were used. Overall $R^2 = 0.6234$ including significant effects only.

Source	SS	df	F	P
log N	1327.0559	1	213.7086	<0.0001
μ	87.6338	1	14.1125	0.0002
$\mu \times \log N$	25.2064	1	4.0592	0.0457
GLR	59.8095	1	9.6317	0.0023
GLR $\times \log N$	57.8781	1	9.3207	0.0027
$\mu \times$GLR	25.1261	1	4.0463	0.0460
$\mu \times$GLR $\times \log N$	12.3659	1	1.9914	0.1602
Error	943.8670	152		

Elena SF *et al.* 2007 *Evolution* **61**, 666-74

Sensitivity to mutational effects depend on μN

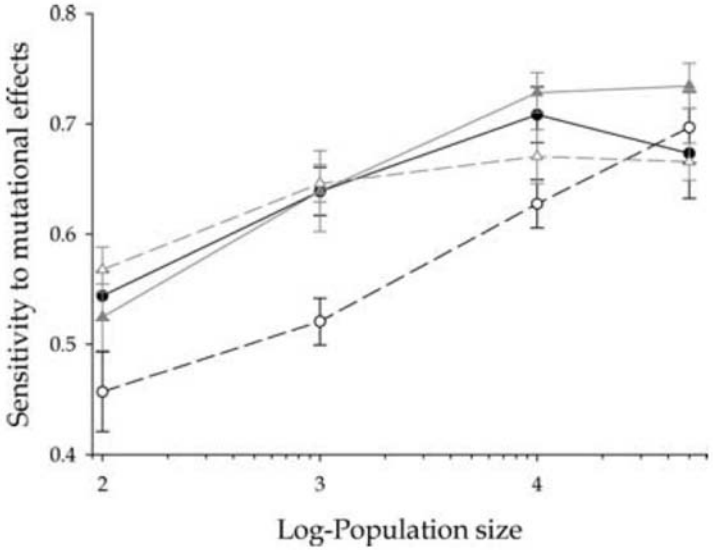


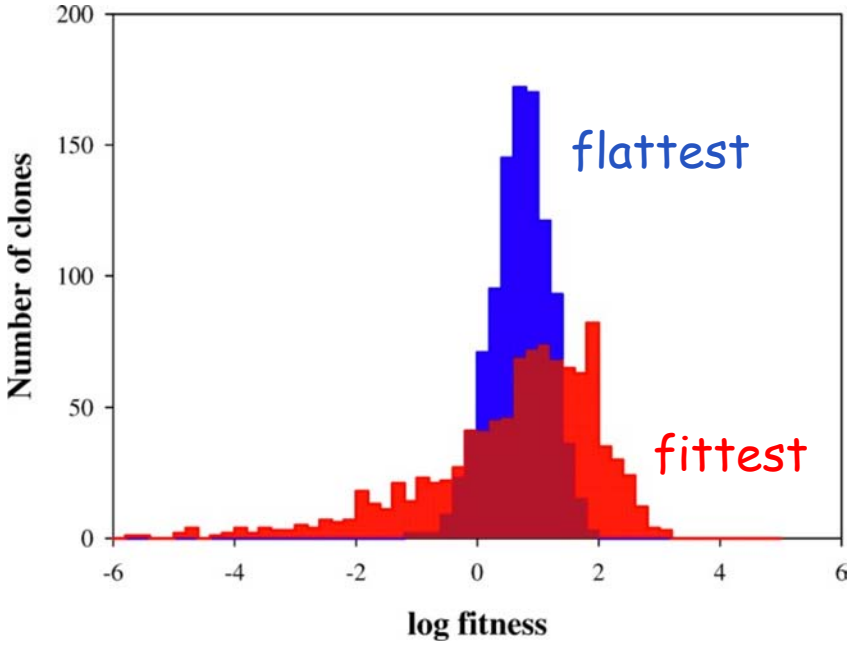
Figure 2. Effects of population size, mutation rate, and genome-length on the mutational sensitivity of the most abundant genotypes isolated at the end of the evolution experiments. Symbols and lines are as described in Figure 1.

Table 2. Analysis of covariance of the mutational sensitivity data shown in Figure 2. See Table 1 for further details. Overall $R^2 = 0.4381$ including significant effects only.

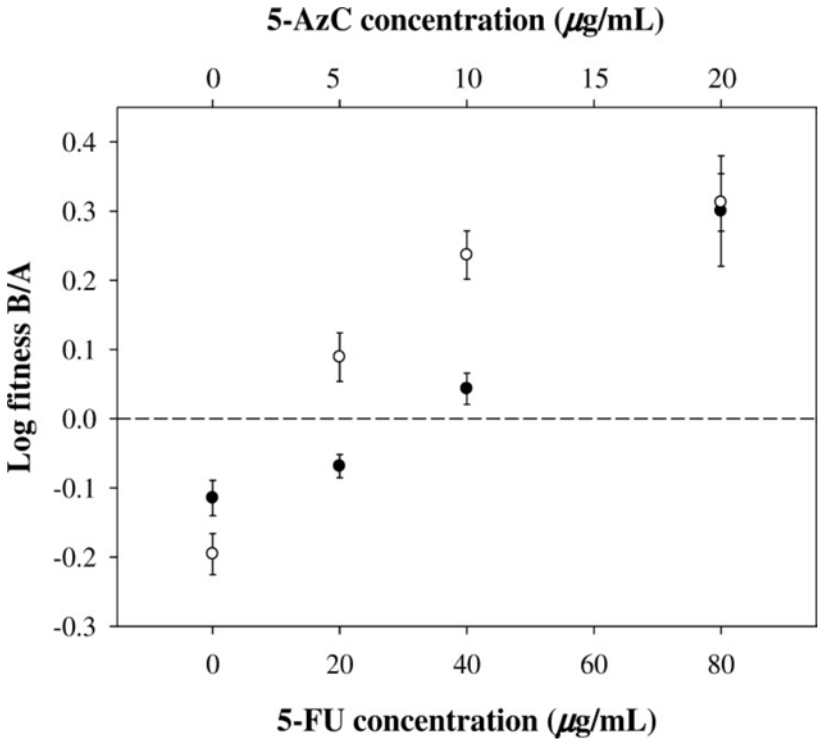
Source	SS	df	F	P
log N	0.7034	1	99.8090	<0.0001
μ	0.0032	1	0.4473	0.5046
$\mu \times \log N$	0.0005	1	0.0643	0.8002
GLR	0.0228	1	3.2299	0.0743
GLR$\times \log N$	0.0071	1	1.0124	0.3159
$\mu \times$GLR	0.0876	1	12.4303	0.0006
$\mu \times$GLR$\times \log N$	0.0703	1	9.9780	0.0019
Error	1.0713	152		

Elena SF *et al.* 2007 *Evolution* **61**, 666-74

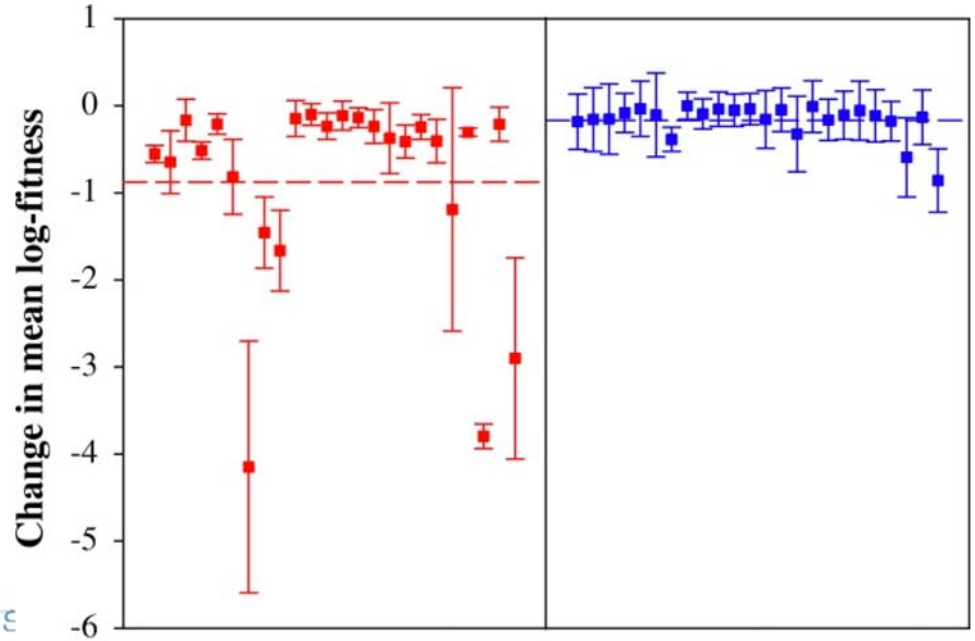
VSV



Flattest wins the competition at increased mutation rate



flattest is more robust to the accumulation of deleterios mutations



Sanjuán R *et al.* 2007 *PLoS Genet.* **3**, e93

SANTA FE INSTITUTE

Evolvability: exploring the fitness landscape

The evolution of optimal mutation rates

U_{opt} = 4.641 mutations/genome/replication

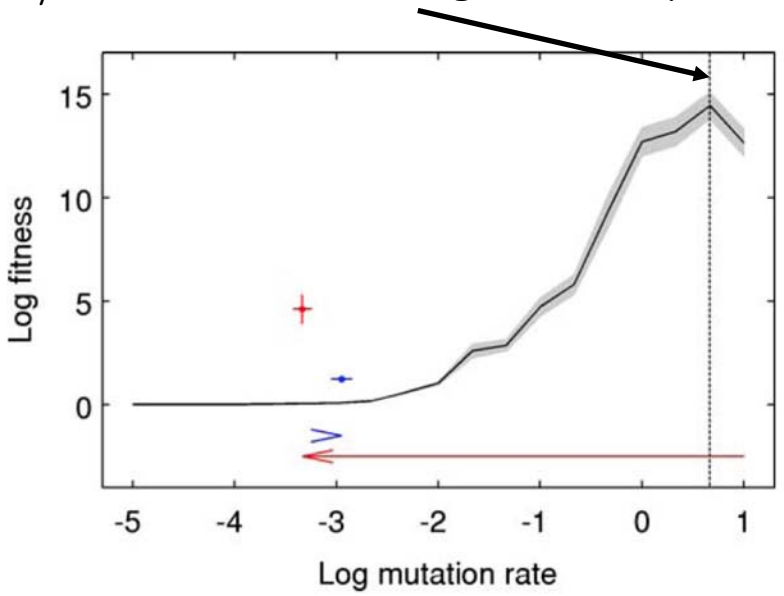


Figure 1. Evolution of suboptimal mutation rates on a complex fitness landscape. Fitness is shown as a function of the genomic mutation rate. The solid line shows mean fitness of the final population, itself averaged over 50 runs, for 15 different static mutation rates ($U = 10^{-5}$, 10^{-4} and from 10^{-3} to 10 at 1/3 $\log_{10}$ intervals). The shaded area represents ±1 s.e.m. The optimal mutation rate—the rate that maximized final fitness—was $U_{opt} \approx 4.641$ (vertical dashed line). The two colored points show the mean fitness and mutation rate of the final population, averaged over 50 runs, in experiments where mutation rates freely evolved with starting values of either 10 (red) or 10^{-3} (blue) (error bars represent ±1 s.e.m). Evolved mutation rates and fitness values were both orders of magnitude lower than those observed in the experiment with U_{opt}.

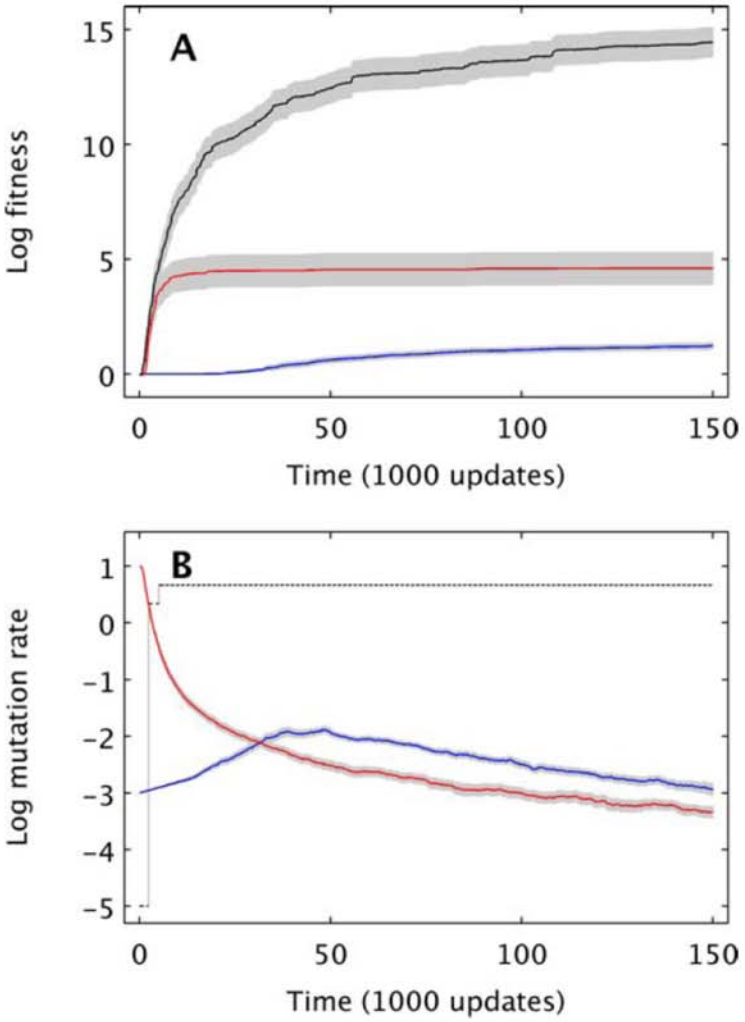


Figure 2. Evolutionary trajectories for fitness and mutation rate on a complex fitness landscape. (A) Evolution of average log-fitness ±1 s.e.m. for treatments with the mutation rate fixed at $U_{opt} = 4.641$ (black) and for treatments with variable mutation rates starting at either 10 (red) or 10^{-3} (blue). (B) Evolution of average log genomic mutation rate ±1 s.e.m. for treatments with variable mutation rates starting at either 10 (red) or 10^{-3} (blue). The black line indicates the mutation rate that had produced the highest average fitness for that time point.

Clune J *et al.* 2008 *PLoS Comput. Biol.* **4**, e1000187

The ruggedness of the fitness landscape determines the value of mutation rate

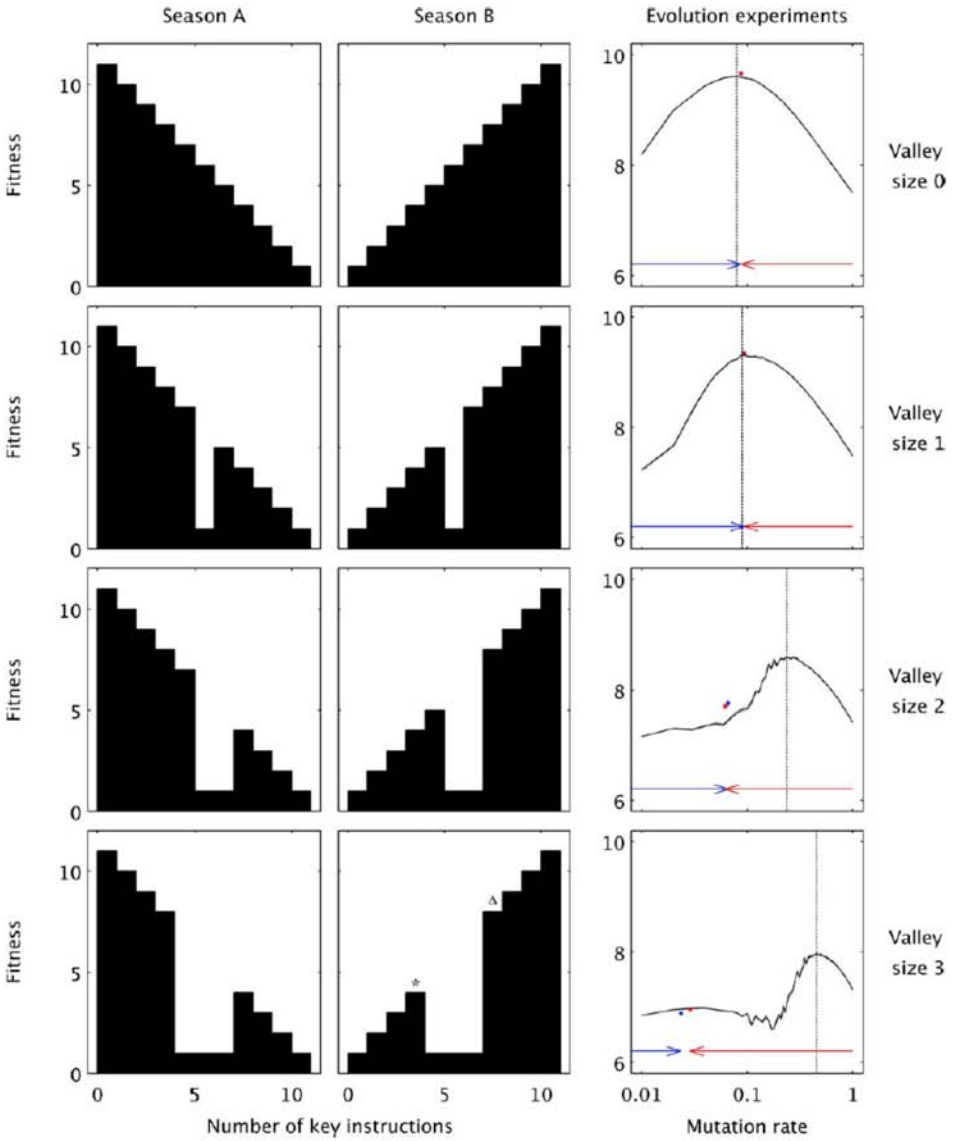


Figure 3. Evolution of mutation rates on simple fitness landscapes with different ruggedness. Here, fitness depended solely on the match between the environment and the number of a key instruction that organisms had in their genomes. In season A (left column) the key instruction was deleterious while it was beneficial in season B (center column). Rugged fitness landscapes with maladaptive valleys (rows 2–4) were introduced by setting the fitness of organisms with intermediate numbers of the key instruction to the minimum fitness level of one. The right-most column shows the results of evolution experiments under each of these selective regimes. Final fitness is shown as a function of genomic mutation rate for both static and dynamic mutation rates. The solid black line represents the average of the mean fitness across 10 runs for each of 100 different static mutation rates ranging from $U = 0.01$ to 1 in increments of 0.01. The two colored points represent the mean fitness and mutation rate,

Clune J *et al.* 2008 *PLoS Comput. Biol.* **4**, e1000187

Selection favors suboptimal mutation rates because they are advantageous in the short term

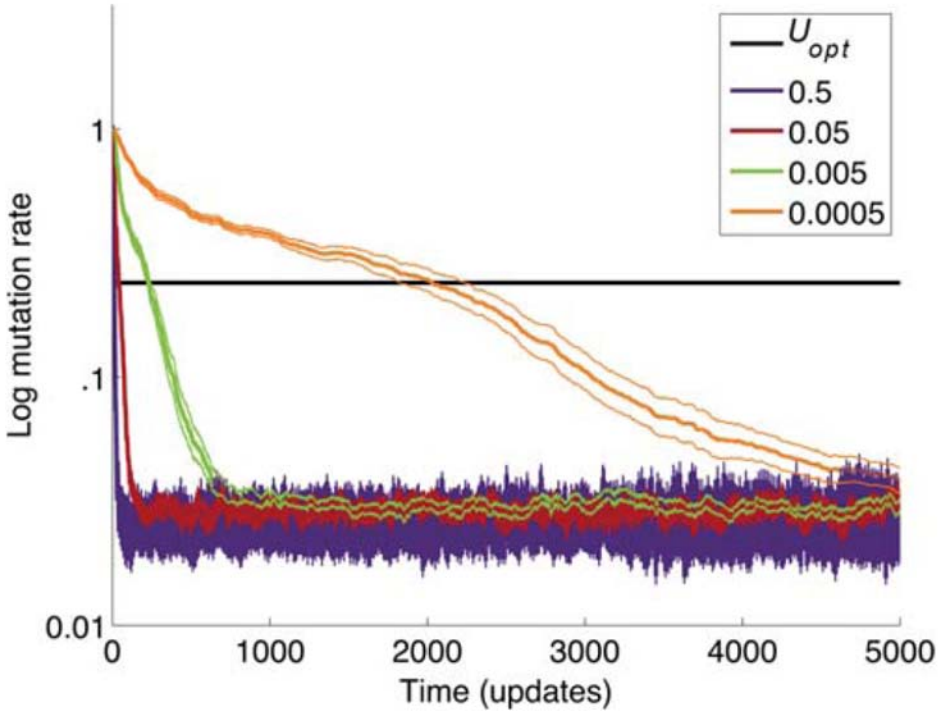


Figure 4. Evolutionarily stable mutation rate does not depend on the frequency with which the mutation rate changes (Π). The evolution of mutation rates in the explicit fitness landscape with a valley size of three is shown for several values of Π, as indicated by the colored key. Each curve shows the average of 20 runs; the adjacent bands represent ±1 s.e.m. The value of U_{opt} was determined in previous experiments (see text). The rate of approach toward the evolutionarily stable mutation rate depends on Π, but the equilibrium value itself does not.

Clune J *et al.* 2008 *PLoS Comput. Biol.* **4**, e1000187

Robustness, evolvability and the efficiency exploring the fitness landscape

- ✓ Organisms were evolved for 10^5 updates in a simple environment (NOT, AND, OR, NOR).
- ✓ Three population sizes 10^2, 10^3 and $5 \cdot 10^4$.
- ✓ Three genomic mutation rates 0.02, 0.2 and 2.
- ✓ Total of 180 lineages. Lineages not performing all four tasks were removed.

Table 1: The two organisms with the most extreme genetic robustness values obtained from the preliminary experiment

			Percentage			Evolutionary history	
Organism	**Fitness**	s_d	**Deleterious**	**Neutral**	**Beneficial**	***U***	***N***
F	417.19	-0.746	89.12% (0.833)	10.36%	0.52% (0.005)	0.2	5×10^4
R	466.75	-0.377	56.84% (0.641)	39.88%	3.28% (0.005)	2	5×10^4

The signed average selection coefficient excluding beneficial mutations, $\overline{s_d}$, was used to quantify robustness. Fitness and the percent of deleterious, neutral and beneficial mutations are shown for each organism. Numbers in parenthesis indicate the average magnitude of mutational effects. The genomic mutation rate (*U*) and the population size (*N*) under which each organism evolved in the preliminary experiment are also shown.

- ✓ *F* is the most fragile organism.
- ✓ *R* is the most robust organism.

Elena SF, Sanjuán R. 2008 *BMC Evol. Biol.* **8**, 284

Further adaptation to a simple environment (NAND, ORN, ANDN, XOR)

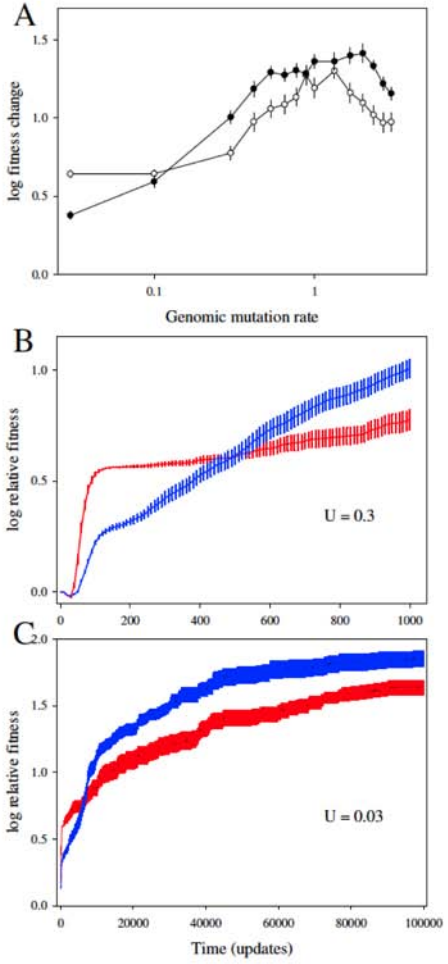


Figure 1
A) Rate of adaptation to a novel 8-task environment as a function of mutation rate, for fragile *F* (white) and robust *R* (black) genotypes. Adaptation was measured as the difference in log fitness of the evolved and ancestral organisms. Fitness was first averaged over all organisms in a population, then log transformed, then averaged over the 50 replicate lineages for each mutation rate. Bars represent standard errors of the mean. B) and C) Fitness trajectories for lineages evolved from *F* (red) and *R* (blue)for two different mutation rates and timescales.

Further adaptation to a very complex environment (77-tasks)

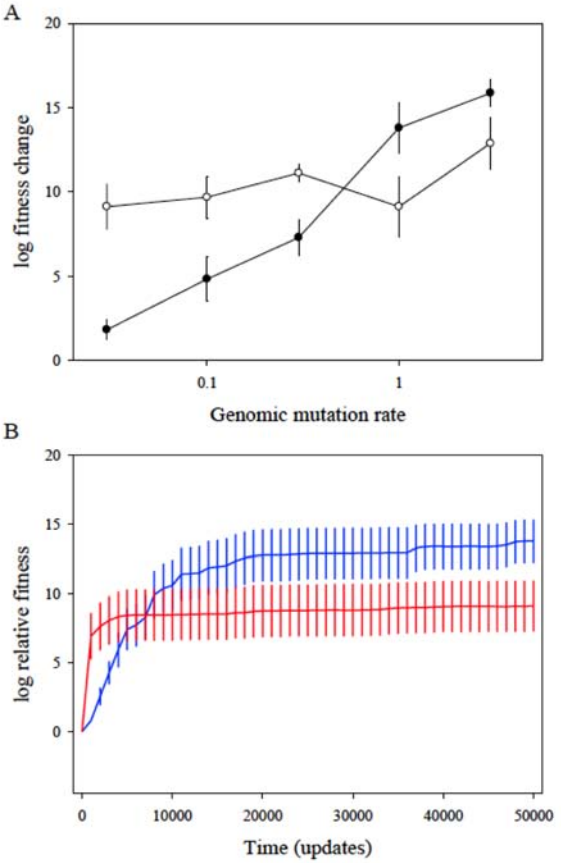


Figure 2
A) Rate of adaptation as a function of mutation rate for a more complex, 77-task, environment (white: fragile genotype *F*; black: robust genotype *R*). B) Fitness versus time for lineages evolved from *F* (red) and *R* (blue) in this environment at a mutation rate *U* = 1. Fitness was calculated in the same manner as in Fig. 1.

Positive association between robustness and evolvability

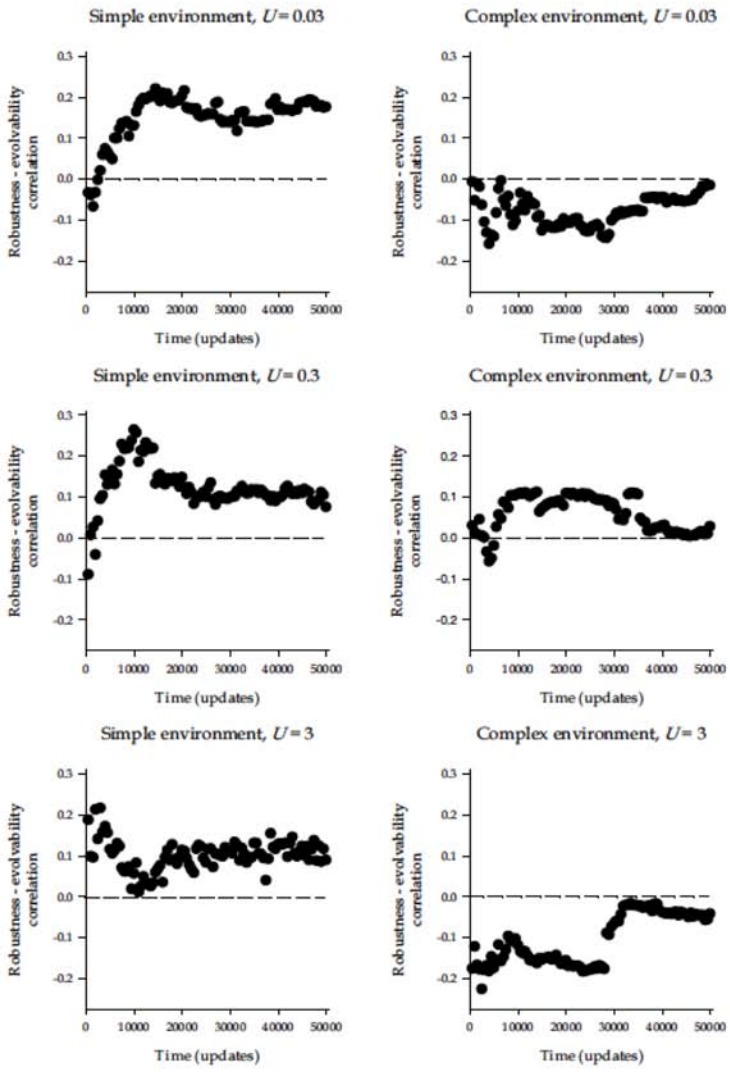
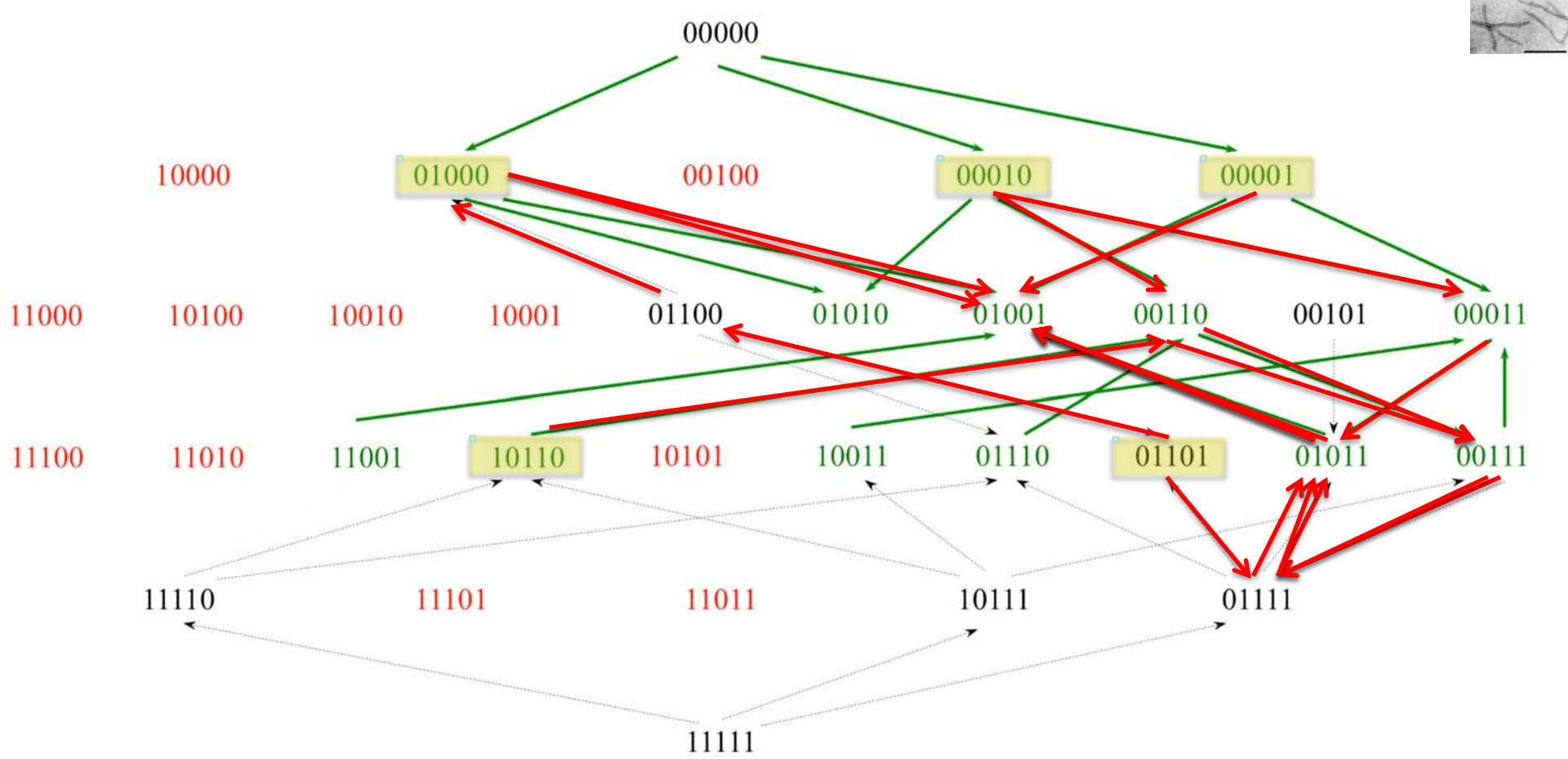


Figure 3
Correlation between robustness and evolvability using all 62 genotypes obtained in the preliminary experiment. The Spearman's correlation between the robustness of the genotype and its rate of adaptation to the two tested environments is displayed versus time. These runs were conducted at the three indicated genomic mutation rates. The null hypothesis of no correlation is indicated by the horizontal dashed lines.

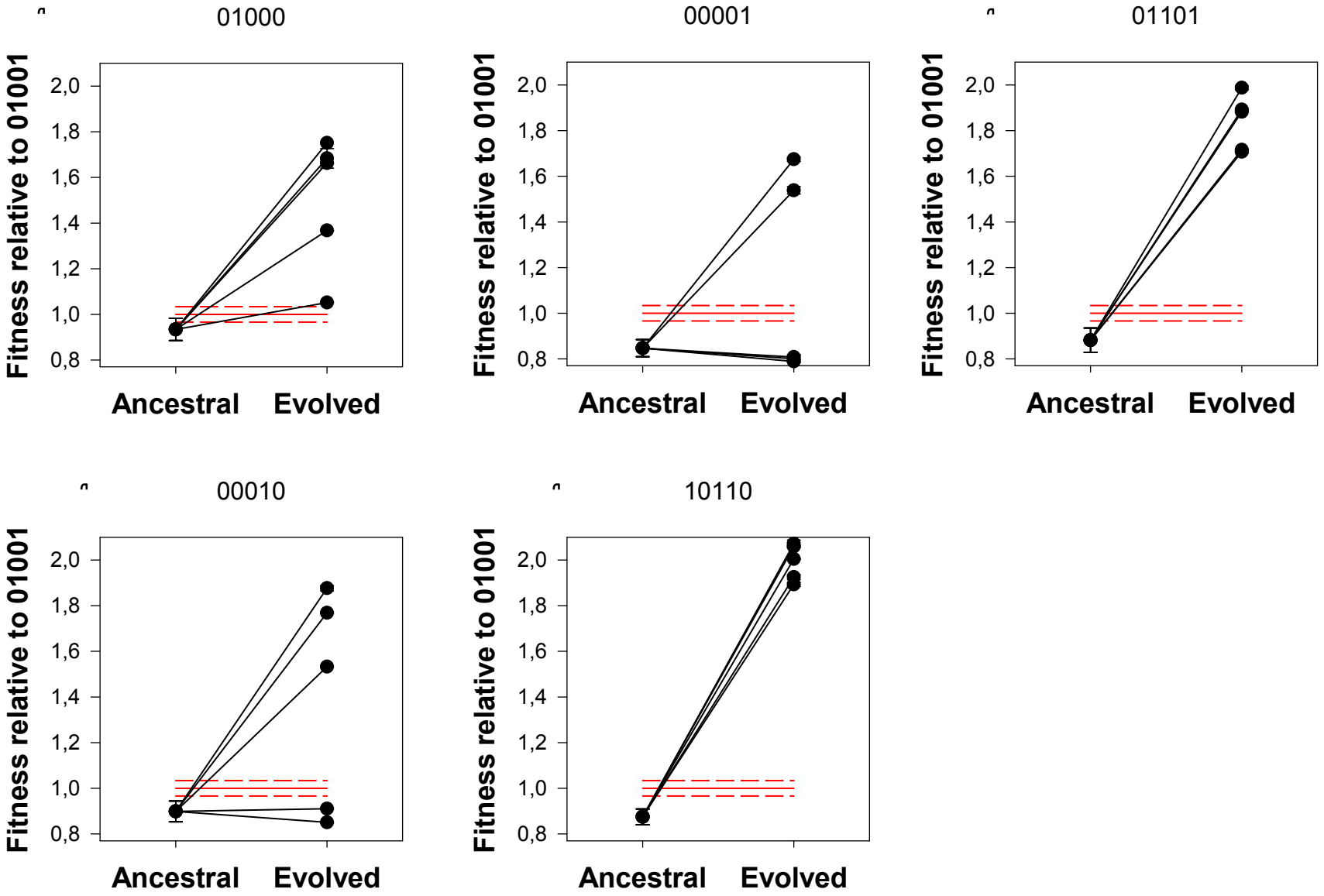
Elena SF, Sanjuán R. 2008 *BMC Evol. Biol.* **8**, 284



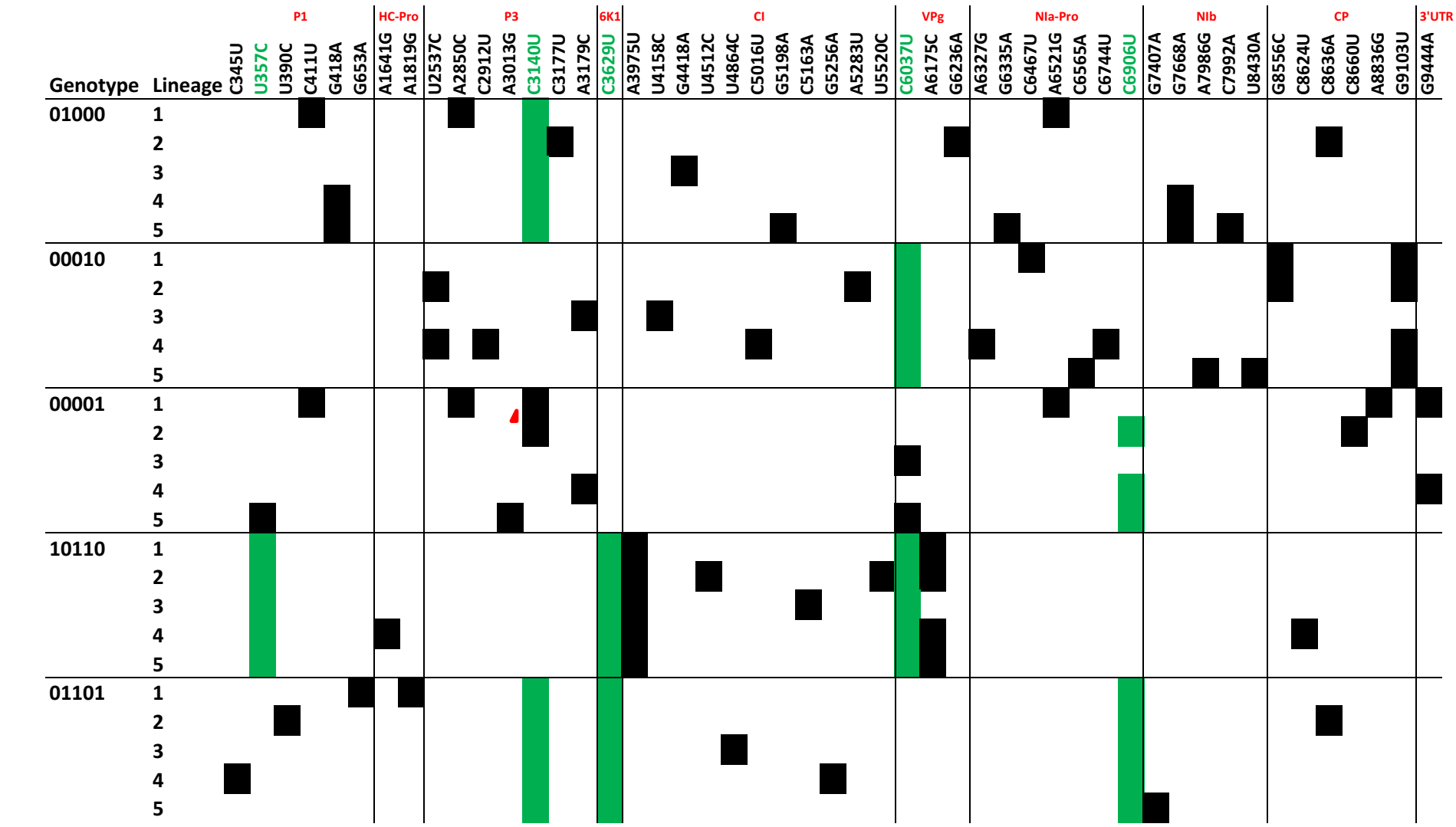
✓ Expected length of the adaptive walk to the optimum genotype:

01000	1 step
00010	3 or 5 steps
00001	1 step
10110	5 steps
01101	3 steps (two paths)

Efficient exploration of distant regions of the landscape



Genomic evolution



Cervera H *et al.* 2016. *Proc. R. Soc. B.* **283**, 20160984

How much evolution depends on past history?

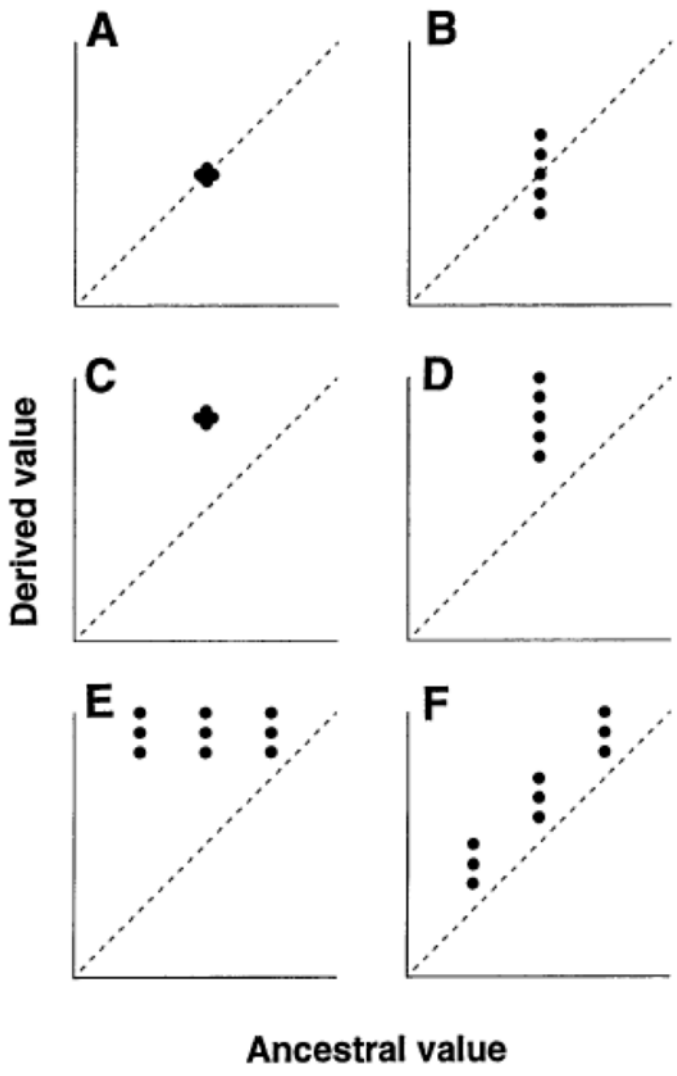
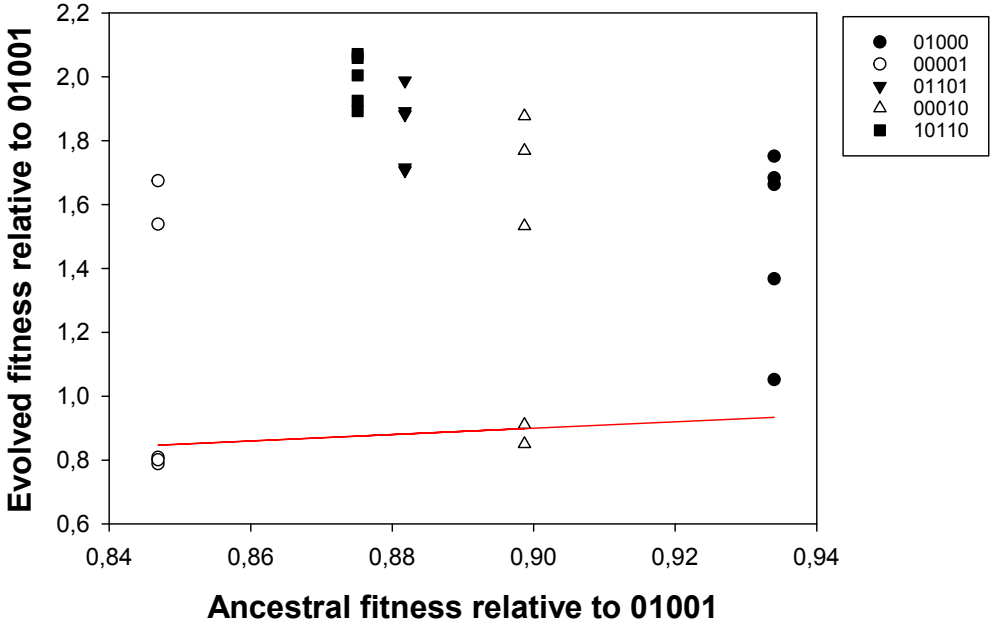
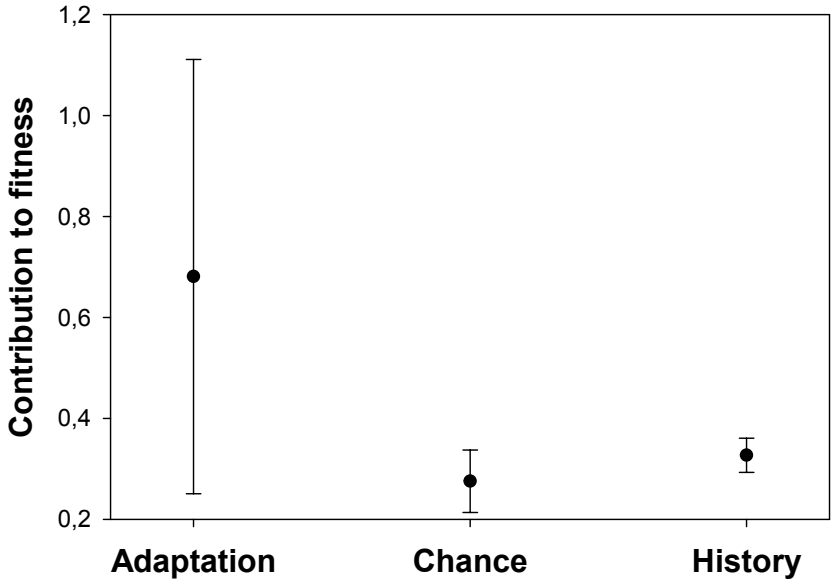


Fig. 1. Schematic representation of effects due to adaptation, chance, and history on evolutionary change and diversification. (**A**) No initial variation and no evolutionary change and hence no effects. (**B**) An effect due to chance only. (**C**) An effect due to adaptation only. (**D**) Effects due to both chance and adaptation. (**E**) An initial effect due to history is eliminated by subsequent effects due to chance and adaptation. (**F**) An initial effect due to history is maintained, with subsequent effects due to chance and adaptation superimposed. See text for further explanation.

Travisano M *et al.* 1995. *Science* **267**, 87-90





- ✓ Observed slope is different from diagonal (*t*-test, *P* = 0.005): initial differences due to history have been erased by subsequent effects due to chance and adaptation.
- ✓ Relative contributions to adaptation: adaptation > chance ~ history

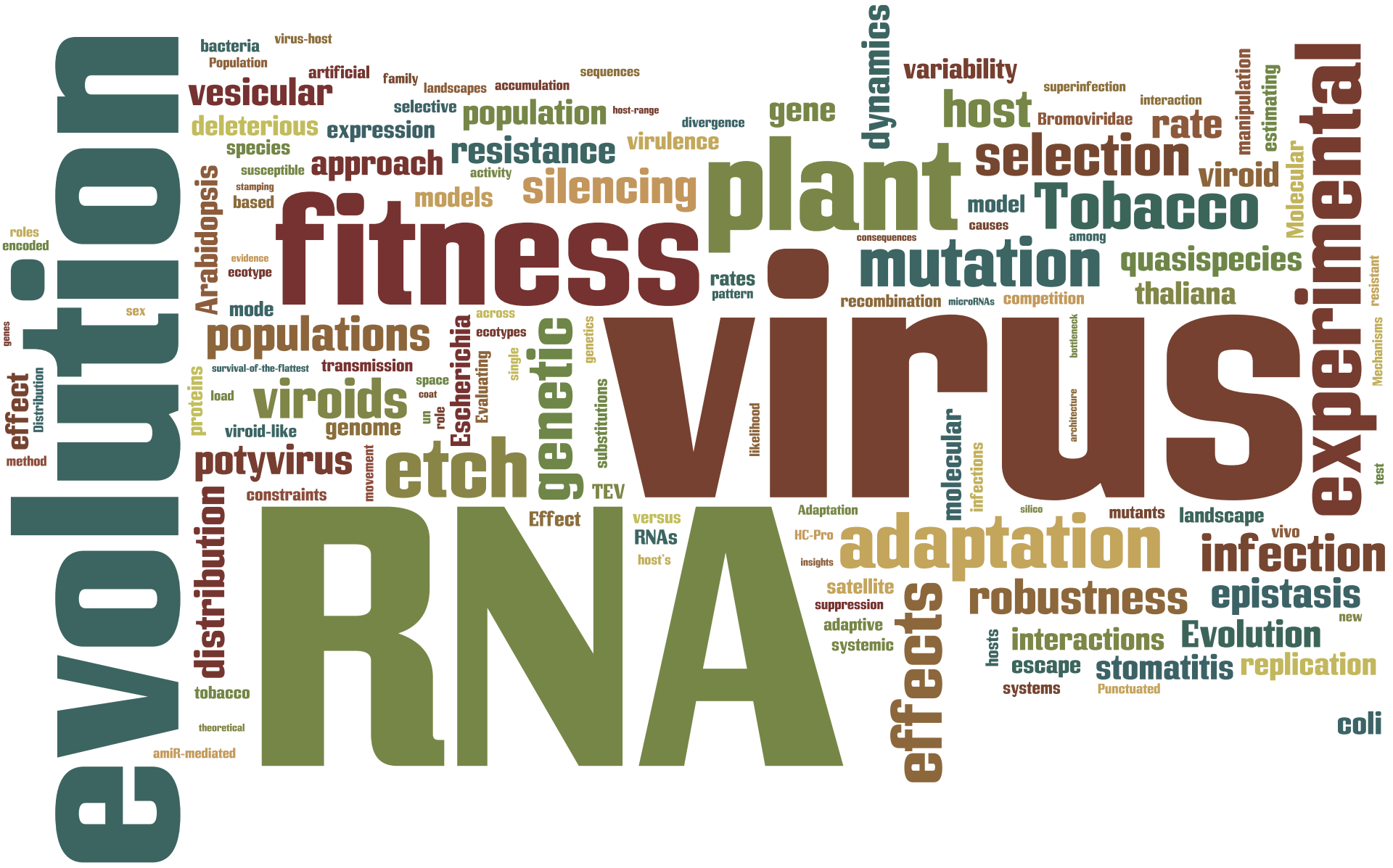
EvolSysVir wetlab

Dr. Héctor Cervera
Dr. Jasna Lalić
Dr. Rafael Sanjuán
Paula Agudo, Francisca de la Iglesia

Colaborators

Prof. Chris Adami (MSU, USA)
Prof. Guillaume Beslon (INRIA-CNRS, FR)
Dr. Jeff Clune (U Wyoming, USA)
Prof. Richard E. Lenski (MSU, USA)
Prof. Charles Ofria (MSU, USA)
Prof. Ricard V. Solé (ICREA-UPF, ES)
Dr. Sergi Valverde (CSIC-UPF, ES)
Prof. Claus O. Wilke (UT Austin, USA)

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
AGENCIA ESTATAL DE INVESTIGACIÓN
SEVENTH FRAMEWORK PROGRAMME
EvoEvo
GENERALITAT VALENCIANA
CONSELLERIA D'EDUCACIÓ, INVESTIGACIÓ, CULTURA I ESPORT
CSIC
Spain National Research Council
EMBO
excellence in life sciences
HIPRA
John Templeton Foundation
HFSP


evolution
RNA
virus
fitness
plant
experimental
adaptation
infection
mutation
selection
Tobacco
host
resistance
silencing
populations
viroids
genetic
etch
potyvirus
distribution
effects
robustness
epistasis
Evolution
interactions
stomatitis
replication
quasispecies
thaliana
Arabidopsis
vesicular
deleterious
expression
population
approach
models
based
dynamics
variability
rate
viroid
Molecular
Escherichia
coli
RNAs
TEV
HC-Pro
satellite
suppression
adaptive
systemic
molecular
infections
likelihood
bottleneck
architecture
genetics
substitutions
movement
constraints
Effect
versus
host's
Adaptation
insights
silico
mutants
landscape
vivo
new
hosts
escape
systems
Punctuated
tobacco
theoretical
amiR-mediated
bacteria
virus-host
Population
artificial
family
landscapes
accumulation
sequences
selective
host-range
divergence
gene
virulence
species
susceptible
activity
stamping
evidence
ecotype
rates
pattern
recombination
microRNAs
competition
consequences
among
model
causes
superinfection
Bromoviridae
interaction
manipulation
estimating
roles
encoded
sex
mode
genes
effect
Distribution
method
proteins
load
viroid-like
genome
survival-of-the-flattest
transmission
space
coat
un
role
Evaluating
single
across
ecotypes
resistant
Mechanisms
test