Can we predict the evolutionary success of a viral strain?

Santiago F. Elena

Evolutionary Systems Virology Group

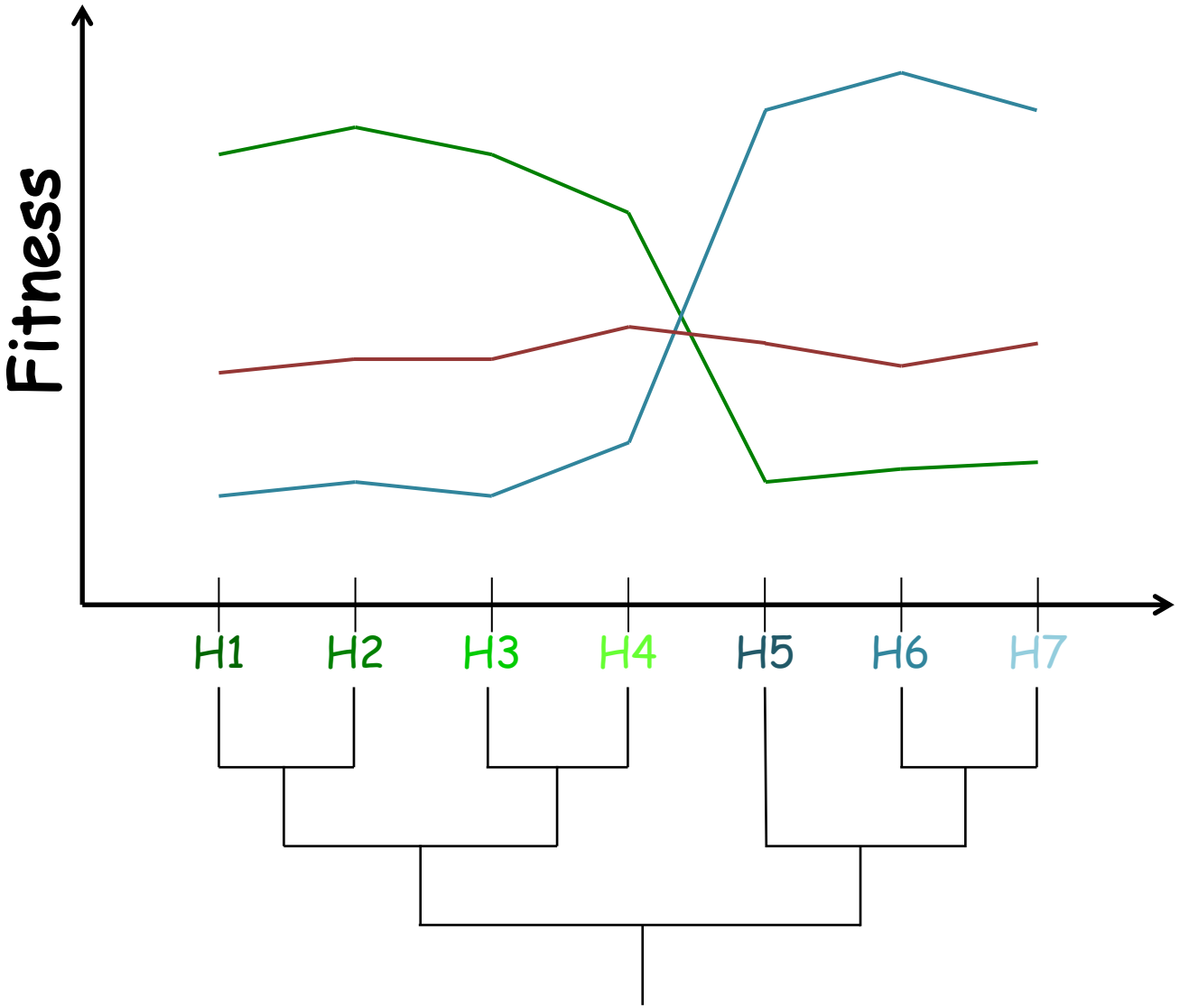
Introduction from reservoir

Successful transmission in the new host

In a perfect world, we would be able of predicting fitness…

$$W = f(G,E) = G + E$$

Fitness tradeoffs across hosts

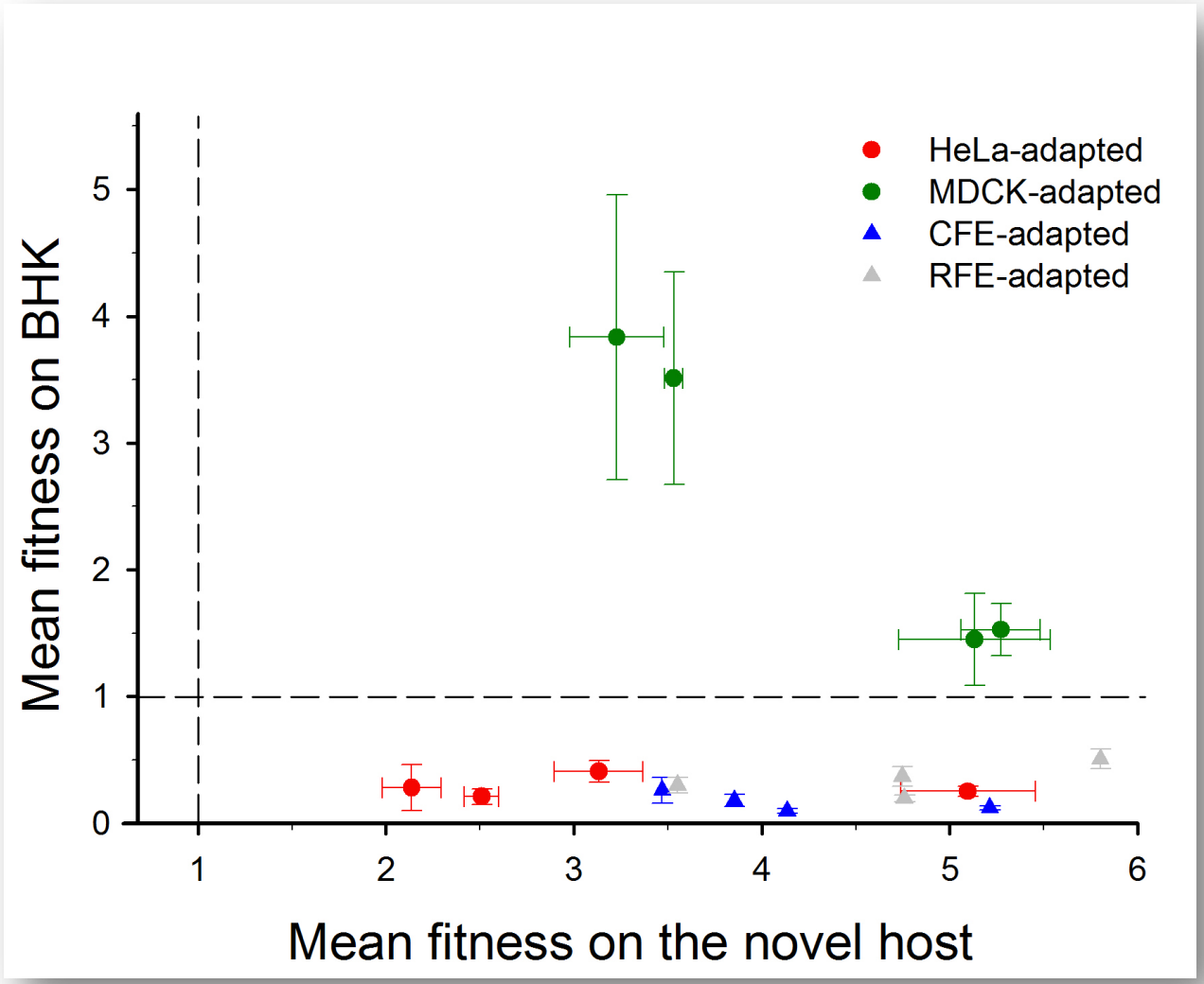


Fitness trade-offs across hosts: causes

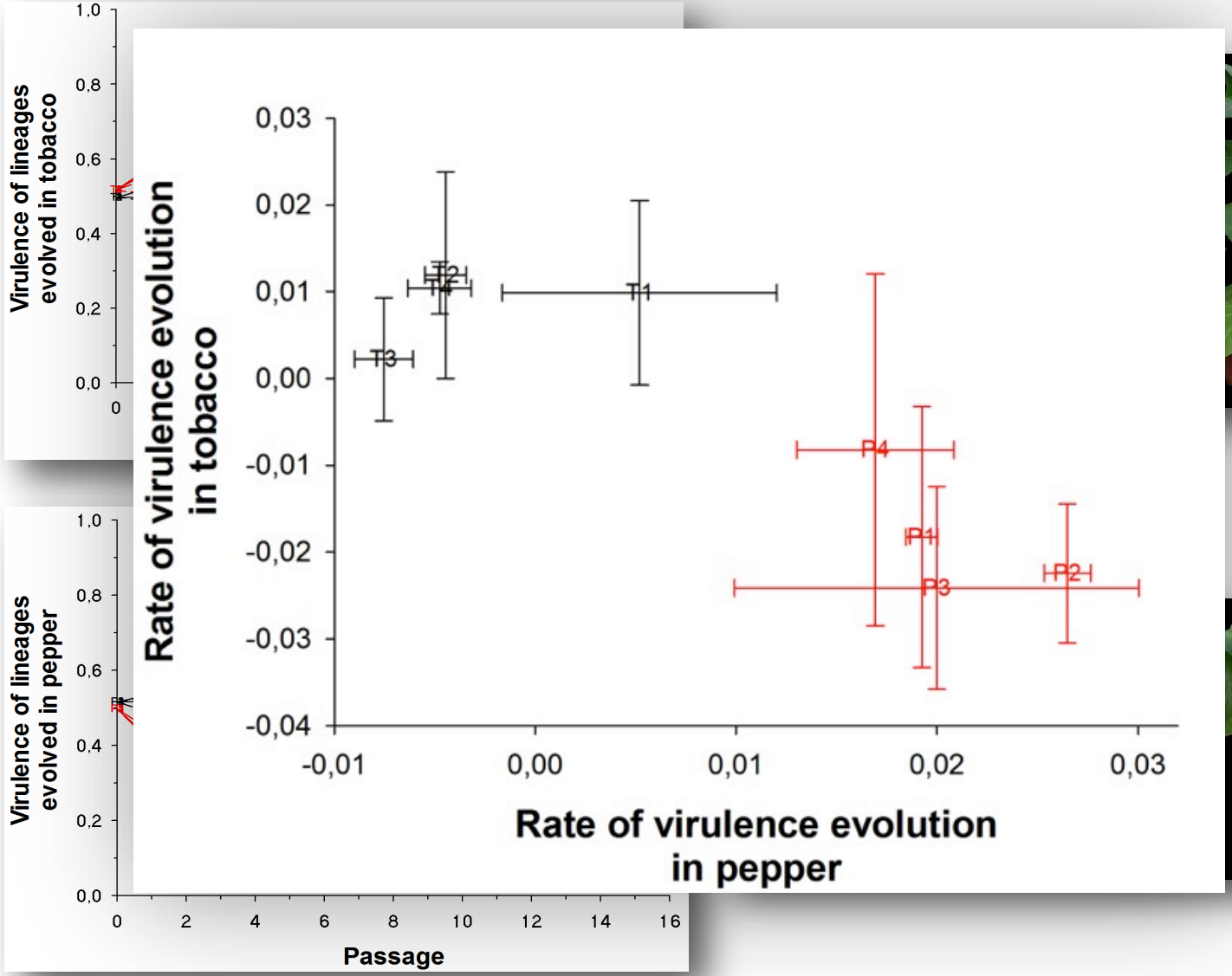
✓ Antagonistic pleiotropy (AP): A particular mutation beneficial in one host is deleterious in another.

Part I
Testing for pleiotropy (G×E)

Evolution in a new single host promotes specialization and pays the cost of host-range expansion.

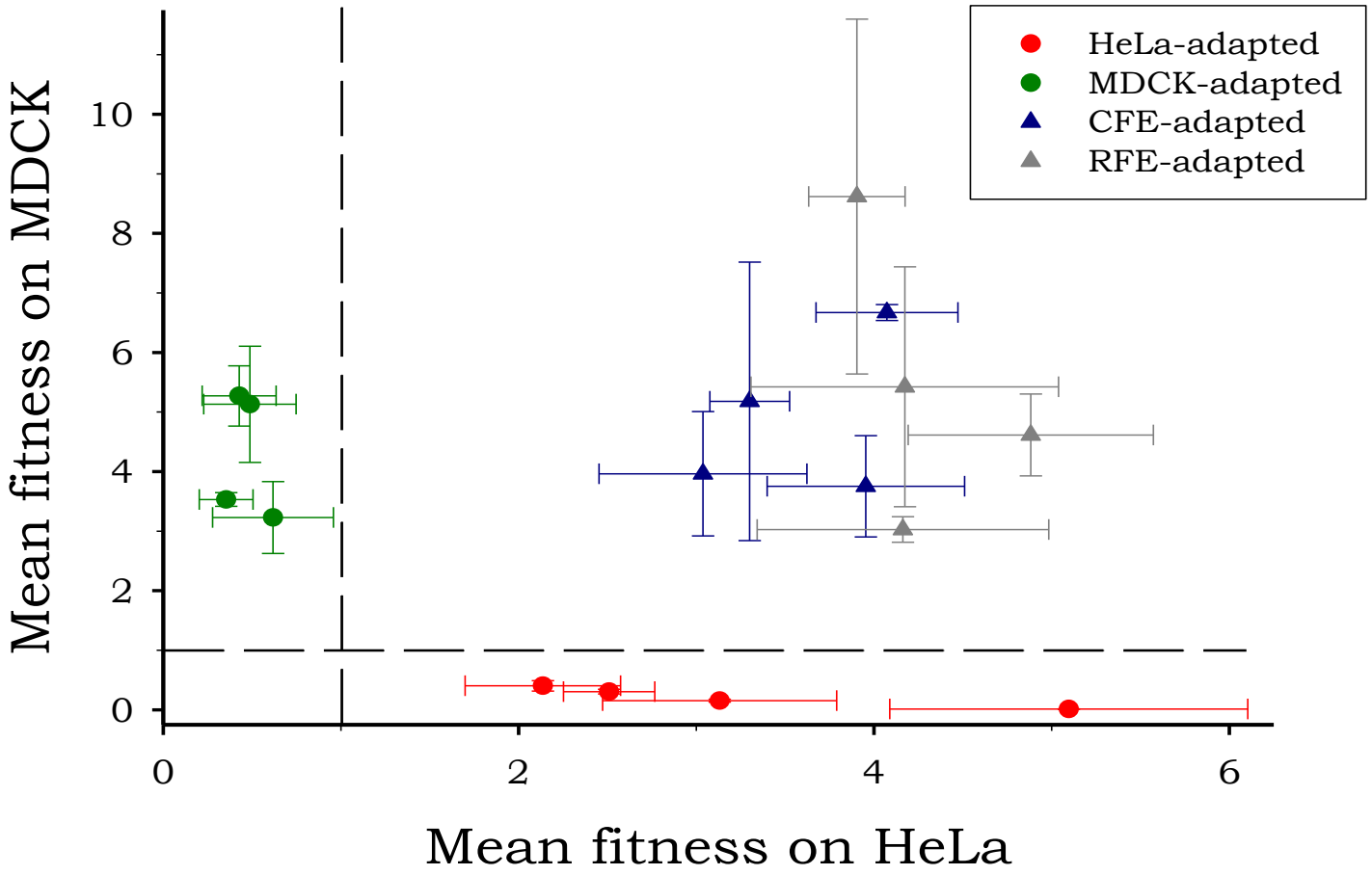


Turner & Elena (2000) *Genetics* **156**, 1465

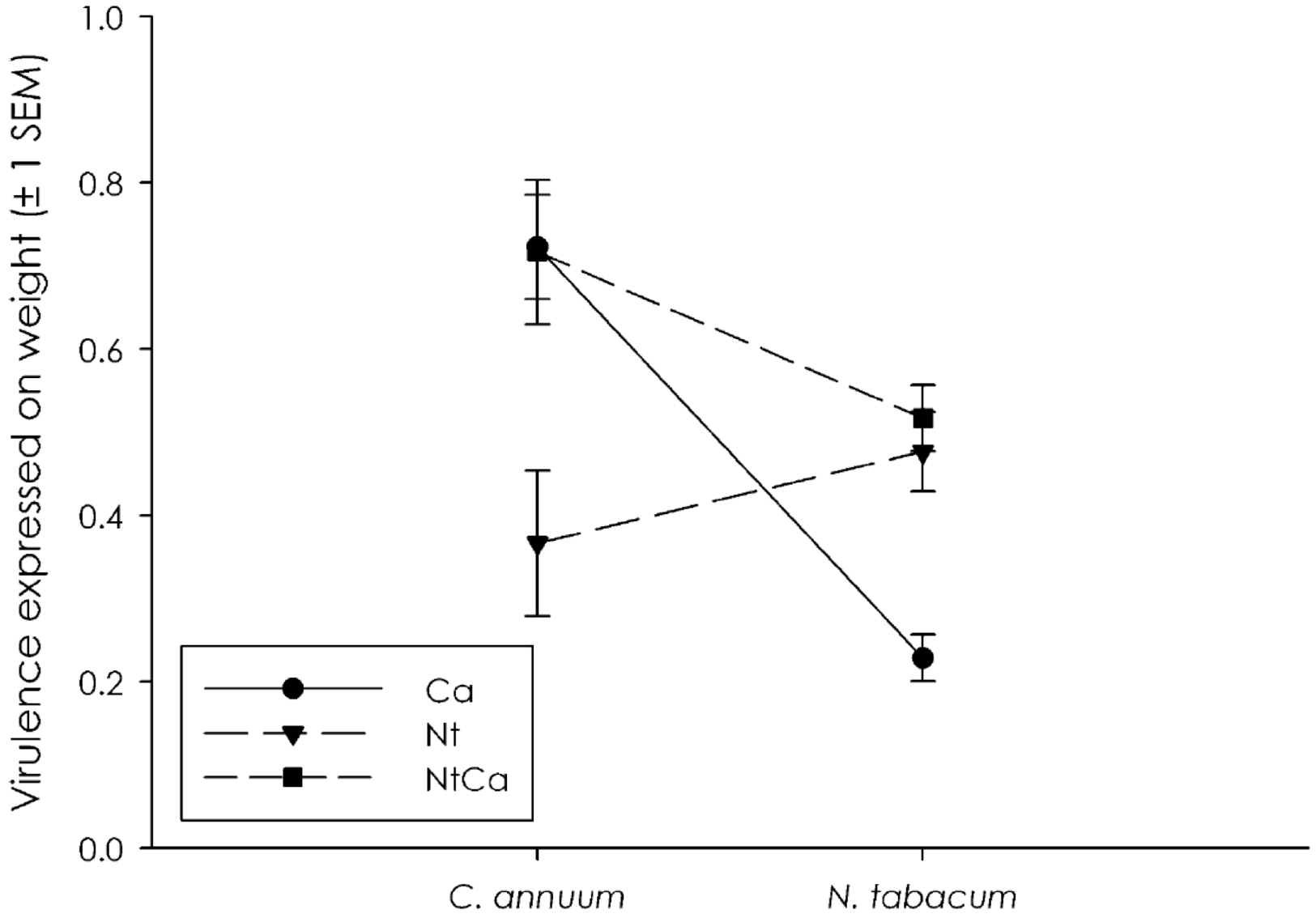


Agudelo-Romero *et al.* (2008) *Infect. Genet. Evol.* **6**, 806
Bedhomme *et al.* (2012) *Mol. Biol. Evol.* **29**, 1481

No-cost generalists evolve under host-switching regimes



Turner & Elena (2000) *Genetics* **156**, 1465



Bedhomme *et al.* (2012) *Mol. Biol. Evol.* **29**, 1481-92

Asterids

Solanaceae

Nicotianoideae

Nicotiana tabacum

Nicotiana benthamiana

Solanoideae

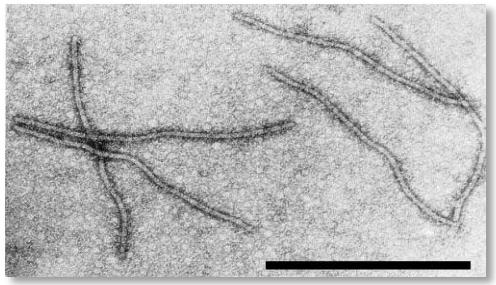
Datura stramonium

Capsicum annuum

Solanum lycopersicum

Asteraceae

Heliantum annuum

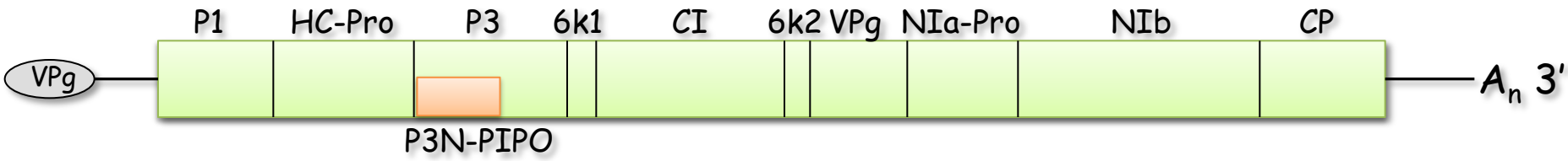


Tobacco etch potyvirus (TEV)

Amaranthaceae

Gomphrena globosa

Spinacea oleracea



4 6+1 8 13 1 6+1 3 8+1 3

1 3 2 2 3

Nonsynonymous changes + stop codons

Synonymous changes

All randomly chosen

Table 2. Parameters describing the DMFE shown in Fig. 1. Frequency of mutations classified as lethal, deleterious, neutral, and beneficial on each host.

	Mean	Median	Std. deviation	Skewness	Kurtosis	Lethal	Deleterious	Neutral	Beneficial
N. tabacum	0.280	0.283	0.016	−1.974	4.608	0	6	14	0
N. benthamiana	0.267	0.277	0.050	−3.949	16.879	0	10	10	0
D. stramonium	0.307	0.322	0.040	−1.566	1.364	2	15	3	0
C. annuum	0.200	0.260	0.116	−1.037	−0.389	0	0	9	11
S. lycopersicum	0.338	0.349	0.029	−0.768	0.062	8	0	2	10
H. annuum	0.026	0.020	0.043	0.527	0.579	0	0	15	5
G. globosa	0.019	0.010	0.041	0.997	0.561	0	0	17	3
S. oleracea	−0.018	−0.039	0.053	1.479	1.915	0	0	17	3

Means are smaller for non-*Solanaceae*: Mann-Whitney *U*-test, $P < 0.001$

Heterogeneity among hosts: χ^2-test, $P < 0.001$

but driven by the difference between *Solanaceae* vs. non-*Solanaceae* : χ^2-test, $P < 0.001$

Table 4. Two generalized lineal models testing the effect of TEV genetic background (*G*), host species (*E*) and their interaction (*G*×*E*). Both variables were treated as random sources.

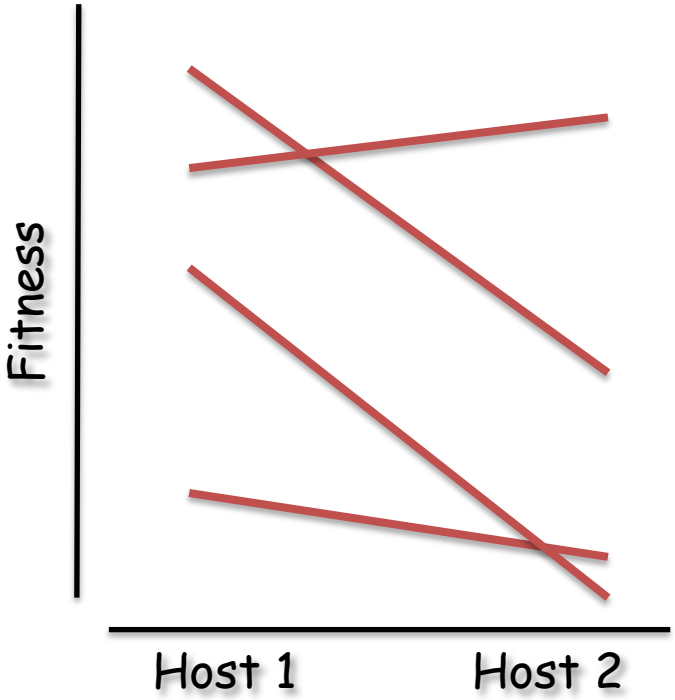
Source of variation	χ^2	d.f.	*P*	Variance component[a]	Percentage of variance[b]
Model I (*AIC*[c] = –2328.299)					
G (TEV genotype)	2783.062	20	< 0.001	4.48×10^{-3}	4.29%
E (Host species)	6467.415	7	< 0.001	2.73×10^{-2}	26.13%
G×*E*	7282.589	140	< 0.001	6.99×10^{-2}	66.82%
Model II (*AIC* = –2412.791)					
G (TEV genotype)	2783.062	20	< 0.001	4.32×10^{-3}	4.17%
Host class	1371.172	1	< 0.001	8.56×10^{-3}	8.25%
E (species within *Host class*)	3177.883	6	< 0.001	1.81×10^{-2}	17.47%
G×*E*	7282.589	140	< 0.001	6.99×10^{-2}	67.33%

[a] Maximum-likelihood estimators.

[b] For *Model I*, computed using a value of error variance equal to 2.88×10^{-3}, which is equivalent to a 2.76% of unexplained variance. For Model II, computed with an error variance 2.88×10^{-3} (2.77%).

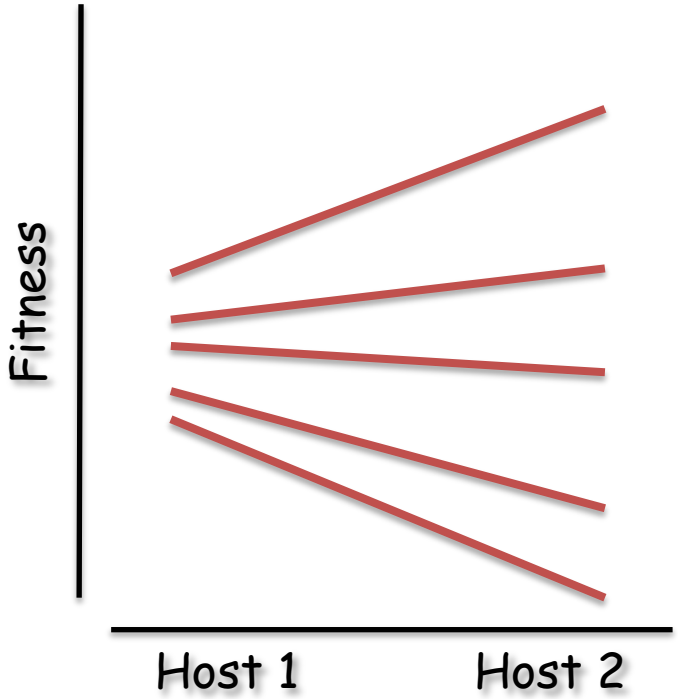
[c] Akaike information criterion.

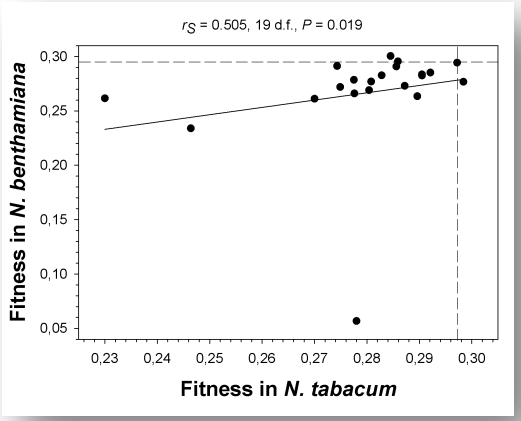
Two hypotheses to explain G×E

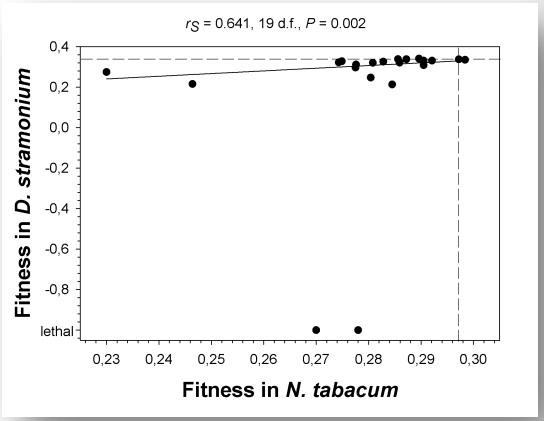


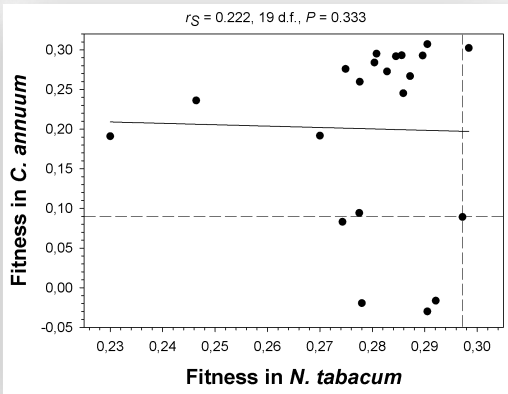
Change in rank order: pleiotropy

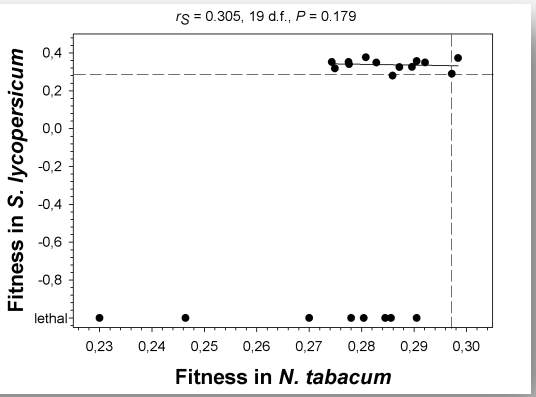
Change in genetic variance for fitness

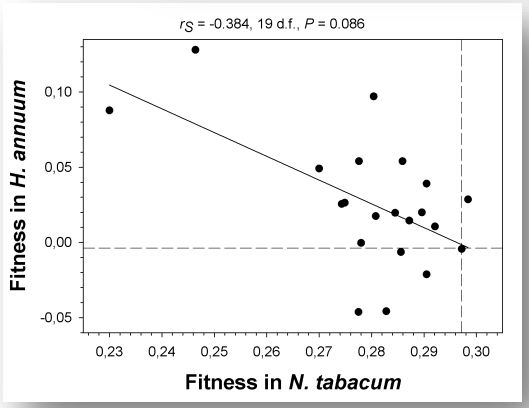


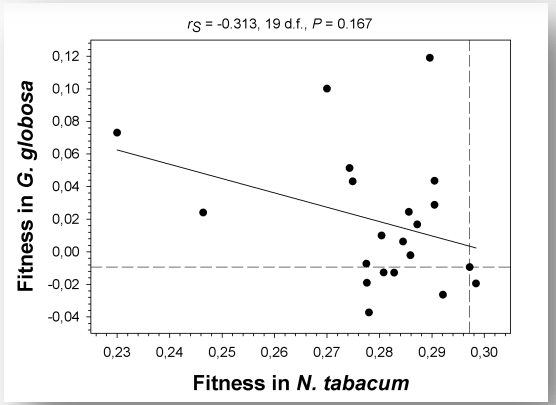


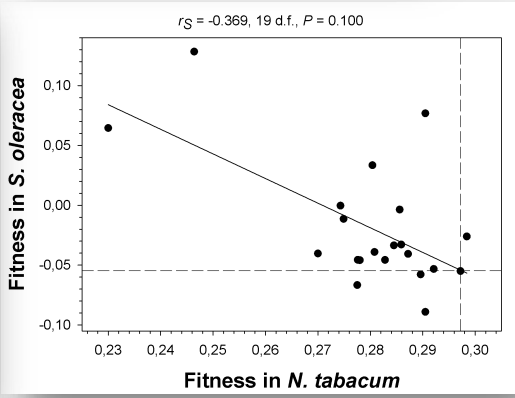












Lalić *et al.* (2011) *PLoS Genet.* **7**, e1002378

- ✓ A non-significant correlation cannot be taken as an evidence of a lack of pleiotropy. One can imagine a situation in which some mutations may have negative pleiotropic effects, some others positive ones and some even being independent on the host.

- ✓ We recorded sign(ΔW) for each mutation on each host. Then we counted the number of cases for which the sign changed between the primary host and each alternative one. Under the null hypothesis of no excess of pleiotropism, sign(ΔW) would distribute evenly.

- ✓ In *N. benthamiana* (2) and *D. stramonium* (4) the number of observed pleiotropic mutations was not significantly larger than expected (Binomial tests: $P = 0.999$ and $P = 0.994$, respectively).

- ✓ The number of mutations that switched signs were significantly larger than expected in all other hosts: 18 in *C. annuum* ($P < 0.001$), 19 in *S. lycopersicum* ($P < 0.001$), 14 in *H. annuus* ($P = 0.058$), 15 in *G. globosa* ($P = 0.021$), and 17 in *S. oleracea* ($P < 0.001$).

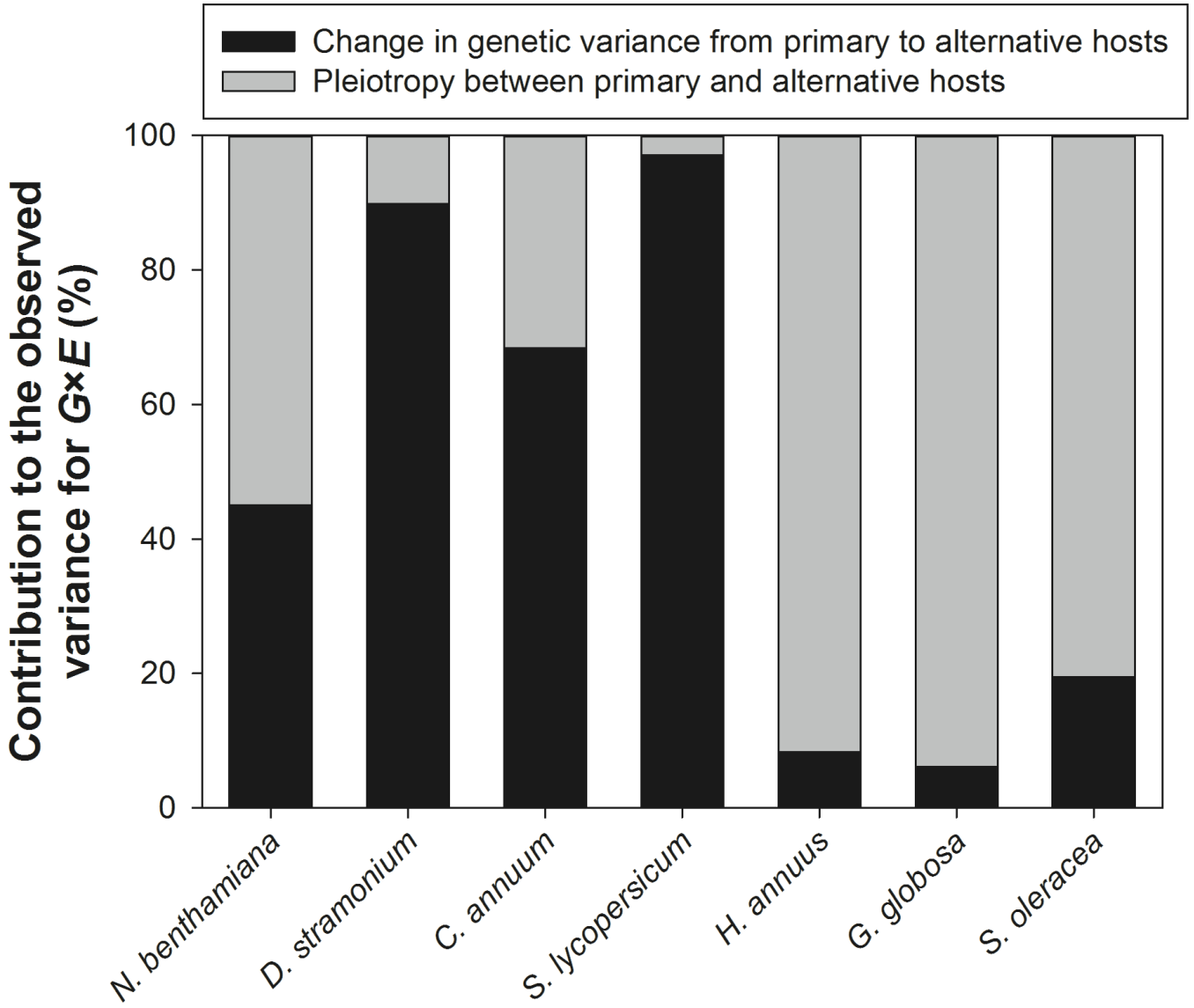
- ✓ Therefore, we confirmed the conclusion of antagonistic pleiotropy as a source of fitness variability across hosts.

- ✓ The relative contribution of these two mechanisms to the observed *G*× *E* can be evaluated using Robertson (1959) decomposition across pairs of hosts (environments):

$$\sigma^2_{G\times E} = \frac{1}{2}\left(\sigma_{G_H} - \sigma_{G_{tobacco}}\right)^2 + \sigma_{G_H}\sigma_{G_{tobacco}}\left(1 - \rho_{G_H G_{tobacco}}\right)$$

Change in magnitude

Pleiotropy



Lalić & S.F. Elena (2013) *Plant Pathol.* **62**, S10

Conclusions Part I

- ✓ The location and shape of DMFE are affected by the host species.
  Location moves towards smaller values as the phylogenetic distance from the reservoir host increases: host are less permissive to infection and support lower virus accumulation.
  Skewness moves towards more positive values as the phylogenetic distance from the host reservoir increases: a larger fraction of mutations are beneficial.

- ✓ Most of observed variability in fitness is explained by $G \times E$ interactions.

- ✓ $G \times E$ interactions have complex origins:
  Unrelated hosts: pleiotropy may drive specialization and balanced polymorphisms.
  Related hosts: changes in genetic variance for fitness affect the balance between selection and drift.

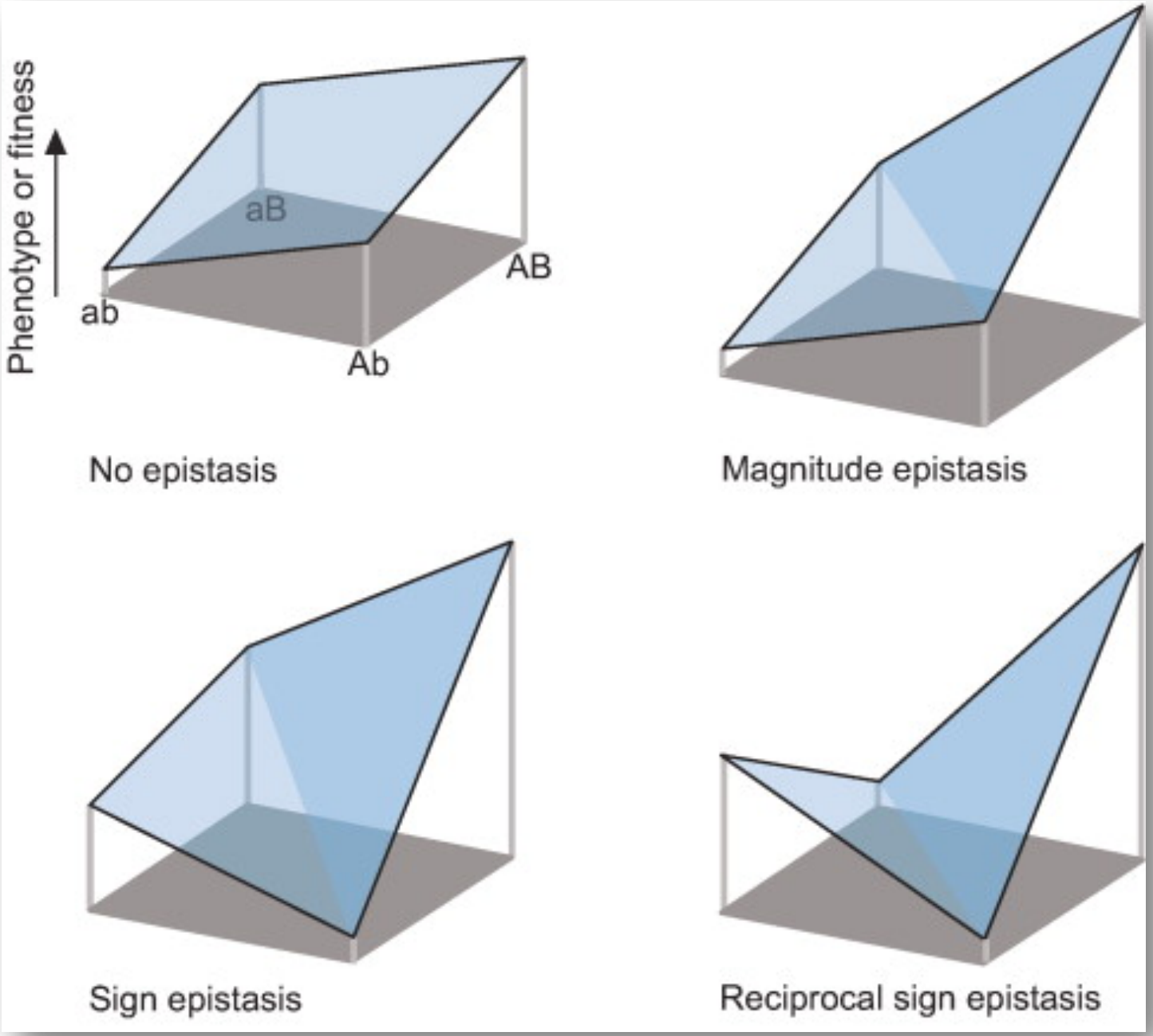
$$W = G + E + G \times E$$

Part II
Testing for epistasis (G×G)

The distribution of *G×G* in the primary host

- ✓ *G×G* aka epistasis, is the interaction between genes or mutations in determining phenotypes.

- ✓ The direction, magnitude and prevalence of epistasis is central to theories seeking to explain the origin of genetic systems, such as sex and recombination, dominance, ploidy, phenotypic plasticity, or robustness, the ruggedness of adaptive landscapes, or attempting to mechanistically explain dynamical biological processes such as the accumulation of mutations in finite populations or speciation by reproductive isolation.


Phenotype or fitness
aB
AB
ab
Ab
No epistasis
Magnitude epistasis
Sign epistasis
Reciprocal sign epistasis

- ✓ We generated a collection of 53 double mutants by combining 20 individual mutations whose deleterious fitness effect had been previously quantified.

- ✓ Mathematical definition of magnitude epistasis:

$$\varepsilon_{xy} = W_{00} W_{xy} - W_{x0} W_{0y}$$

$\varepsilon_{xy} > 0$ positive (antagonistic) epistasis
$\varepsilon_{xy} < 0$ negative (synergistic) epistasis
$\varepsilon_{xy} = 0$ no epistasis (additive)

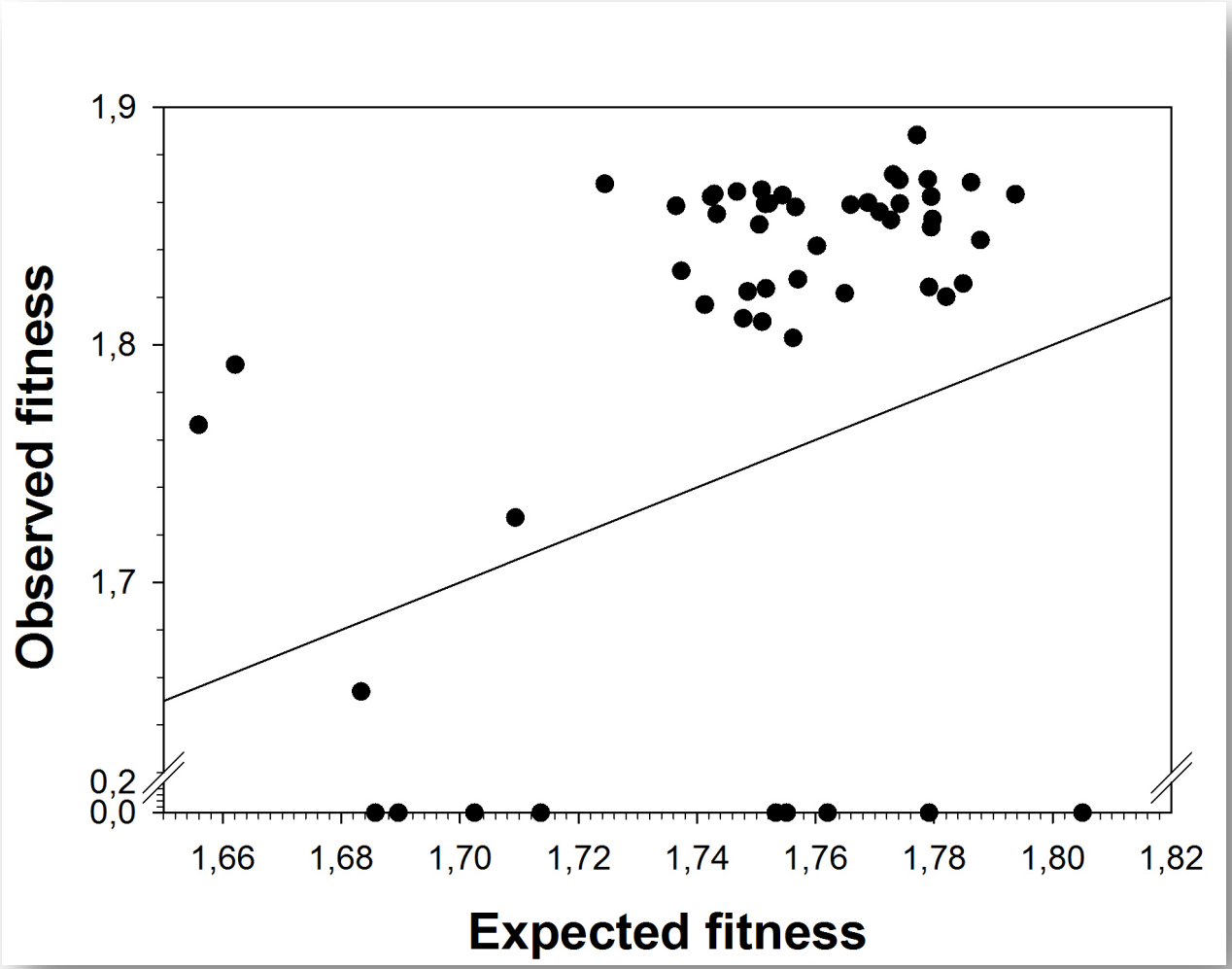
- ✓ Mathematical condition for sign epistasis (F.J. Poelwijk *et al.* (2011) *J. Theor. Biol.* **272**: 141-4):

$$|W_{x0} - W_{00} + W_{xy} - W_{0y}| < |W_{x0} - W_{00}| + |W_{xy} - W_{0y}|$$

- ✓ Additional mathematical condition for reciprocal sign epistasis (F.J. Poelwijk *et al.* (2011) *J. Theor. Biol.* **272**: 141-4):

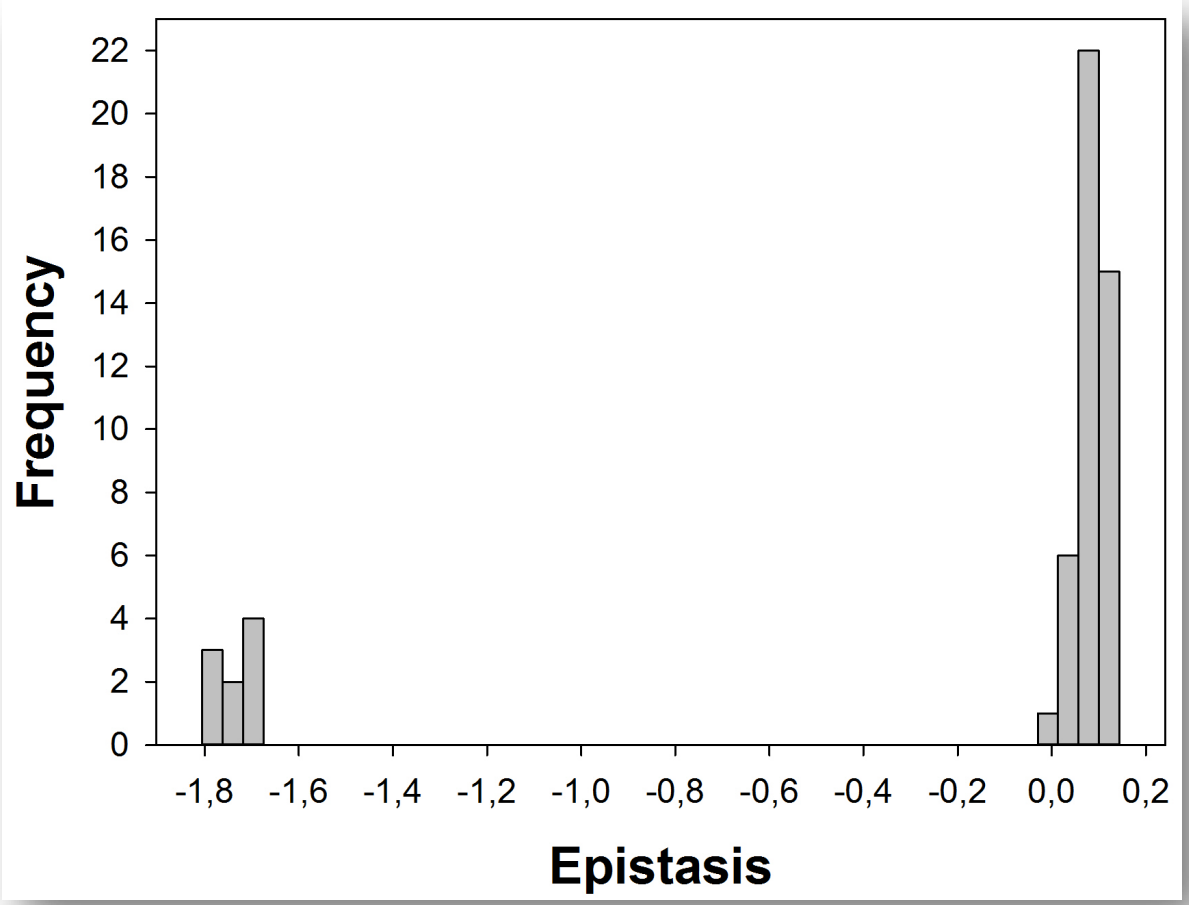
$$|W_{0y} - W_{00} + W_{xy} - W_{x0}| < |W_{0y} - W_{00}| + |W_{xy} - W_{x0}|$$

Epistasis among pairs of deleterious mutations



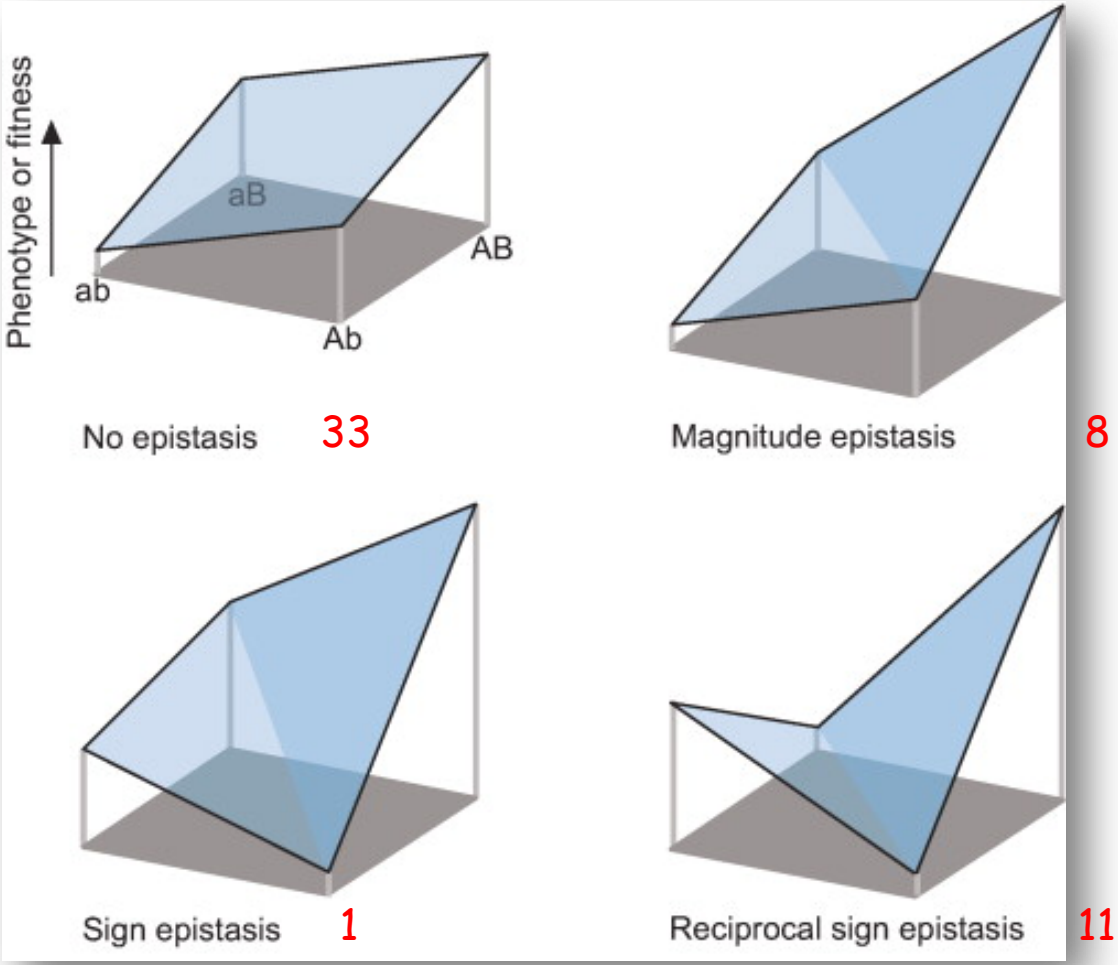
- ✓ 20 significant deviations from the additive expectation (*t*-test, $P < 0.049$).
  9 cases of synthetic lethals (negative epistasis).
  11 cases of positive epistasis.

Statistical properties of the epistasis distribution



- ✓ Without synthetic lethals, $\langle \varepsilon \rangle = -0.224 \pm 0.005$ (t-test, $P < 0.001$).
- ✓ Significant negative skewness ($g_1 = -1.856 \pm 0.328$; $P < 0.001$).
- ✓ Significantly leptokurtic ($g_2 = 2.348 \pm 0.642$, $P = 0.002$).

Pervasive reciprocal sign epistasis



- ✓ 33% less cases of magnitude than of sign epistasis (Binomial test, 1-tailed $P = 0.032$).
- ✓ Over-representation of reciprocal sign epistasis among cases of sign epistasis (Binomial test, $P < 0.001$).

Lalić & S.F. Elena (2012) *Heredity* **109**, 71

Epistasis determines the rate of adaptation

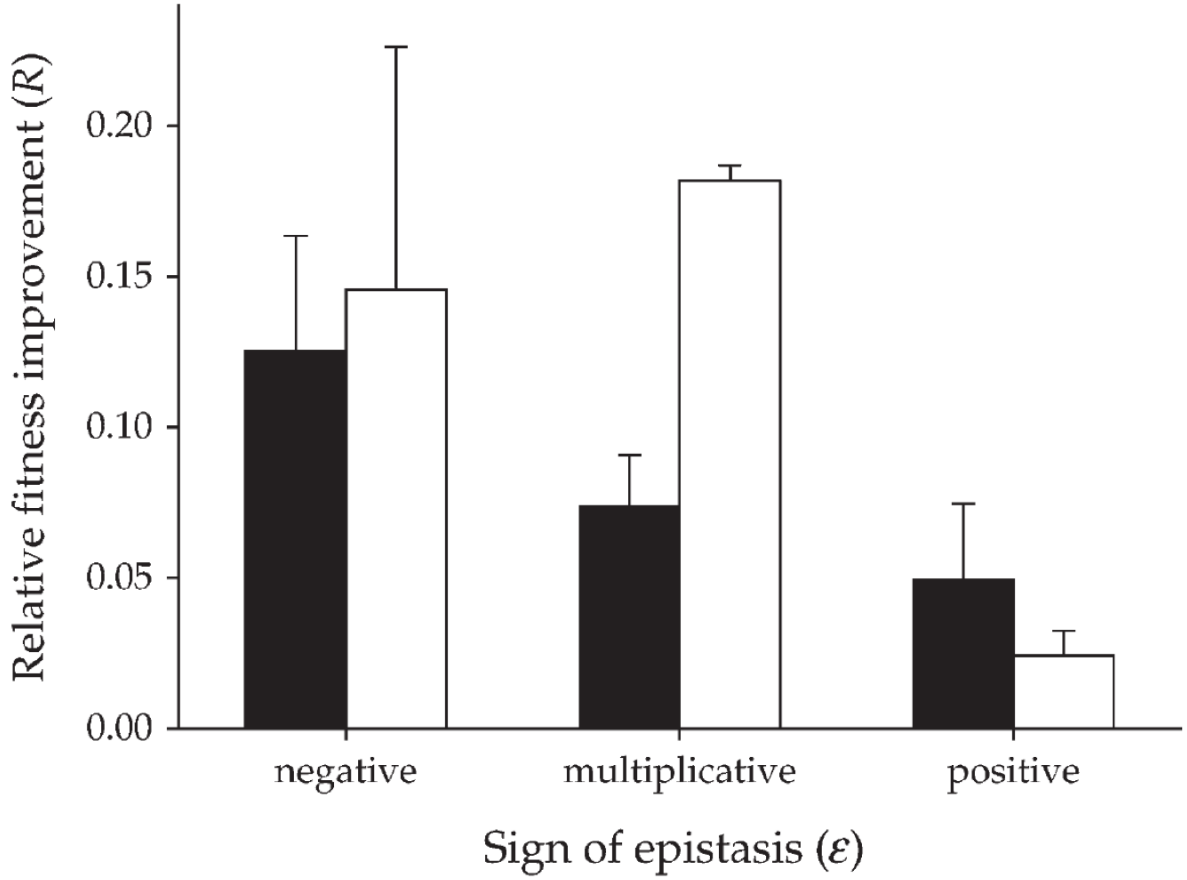


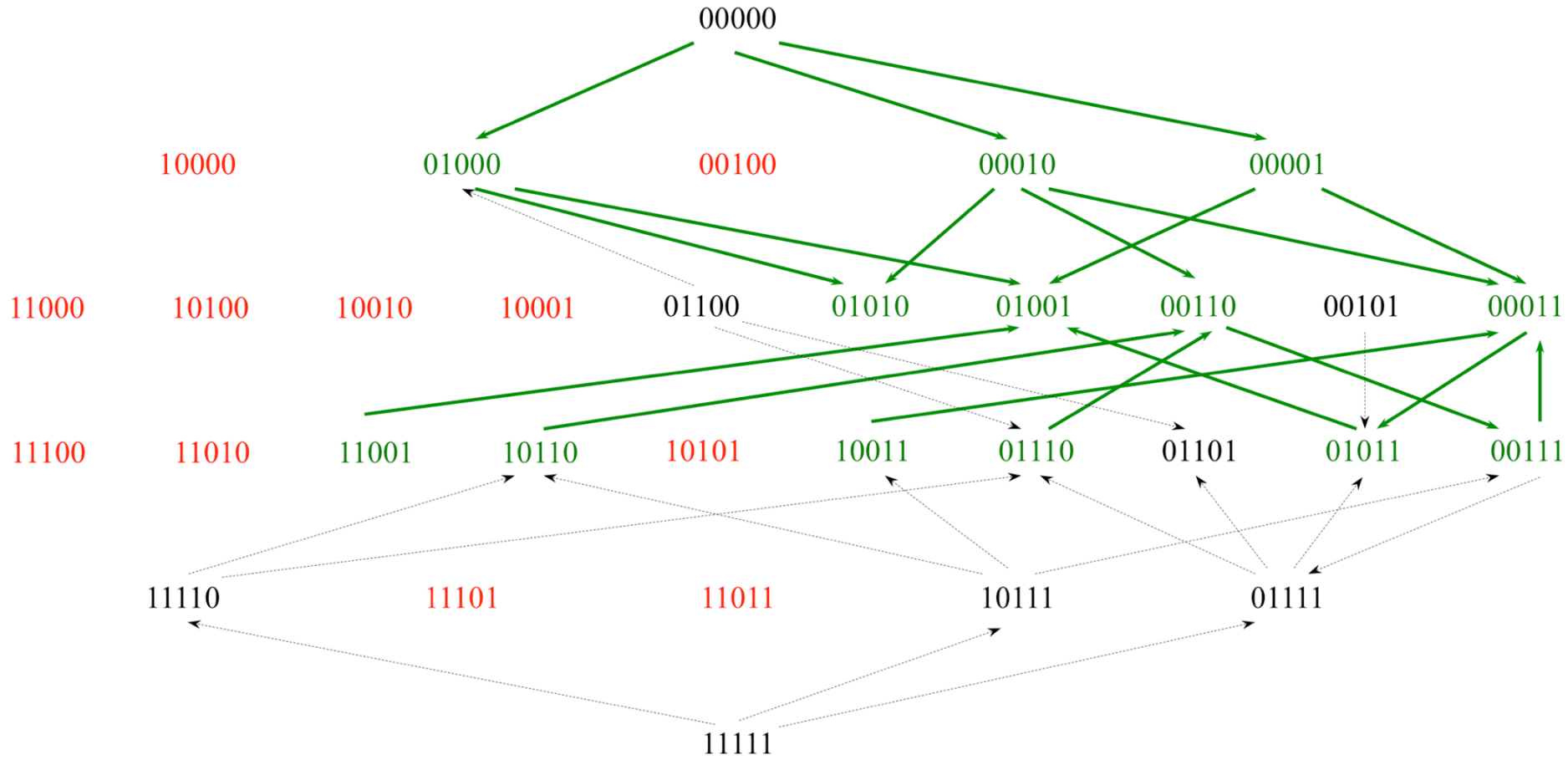
FIGURE 3.—Average fitness improvement as a function of the type of epistasis characteristic of the two mutations carried by the double mutants, for the two effective population sizes. Solid bars show $N_e = 2 \times 10^2$ PFU and open bars show $N_e = 2 \times 10^4$ PFU. Error bars show standard errors of the means.

Sanjuán *et al.* (2005) *Genetics* **170**, 1001

Global topography of a frozen adaptive fitness seascape

Table 1. Mutations fixed in the *A. thaliana*-adapted virus TEV-*At*17. The potential adaptive value of two synonymous substitutions is evaluated by comparing codon usage

						A. thaliana codon usage frequency (per thousand)			
Label	Mutation	Gene	Amino acid change	Wildtype codon	Mutated codon	Wildtype codon	Mutated codon	Fitness	Variance for fitness ($\times 10^{-3}$)
100000	U357C	P1	synonymous	CAU	CAC	13.8	8.7	0.9580	0.6262
010000	C3140U	P3	A1047V	GCG	GUG			1.0540	7.2957
001000	C3629U	6K1	T1210M	ACG	AUG			0.0000	0.0000
000100	C6037U	VPg	L2013F	CUU	UUU			1.0678	7.7630
000010	C6666U	NIa-Pro	synonymous	GGC	GGU	9.2	22.2	nd	nd
000001	C6906U	NIa-Pro	synonymous	UUC	UUU	20.7	21.8	1.0430	4.1553


00000
10000
01000
00100
00010
00001
11000
10100
10010
10001
01100
01010
01001
00110
00101
00011
11100
11010
11001
10110
10101
10011
01110
01101
01011
00111
11110
11101
11011
10111
01111
11111

- ✓ Overall, differences in fitness exist among the 32 genotypes (χ^2 = 706.905, 31 d.f., $P < 0.001$), with 60.16% of observed variance among genotypes being explained by genetic factors.

- ✓ Twenty four genotypes show significant differences with the ancestral TEV: 11 were worse and 13 fitter. The smallest benefit was for 10110 (6.10%) and the largest for 01001, the global optimum (12.01%). Genotype 01011 (11.18%) represents a second local optimum.

- ✓ Application of Poelwijk *et al.* (2011) equations to > 2 mutations: *e.g.*, genotype 10110 could be constructed in three ways: inserting 00010 into 10100, inserting 00100 into 10010, or inserting 10000 into 00110.

- ✓ In this way, we can test 65 possible combinations.

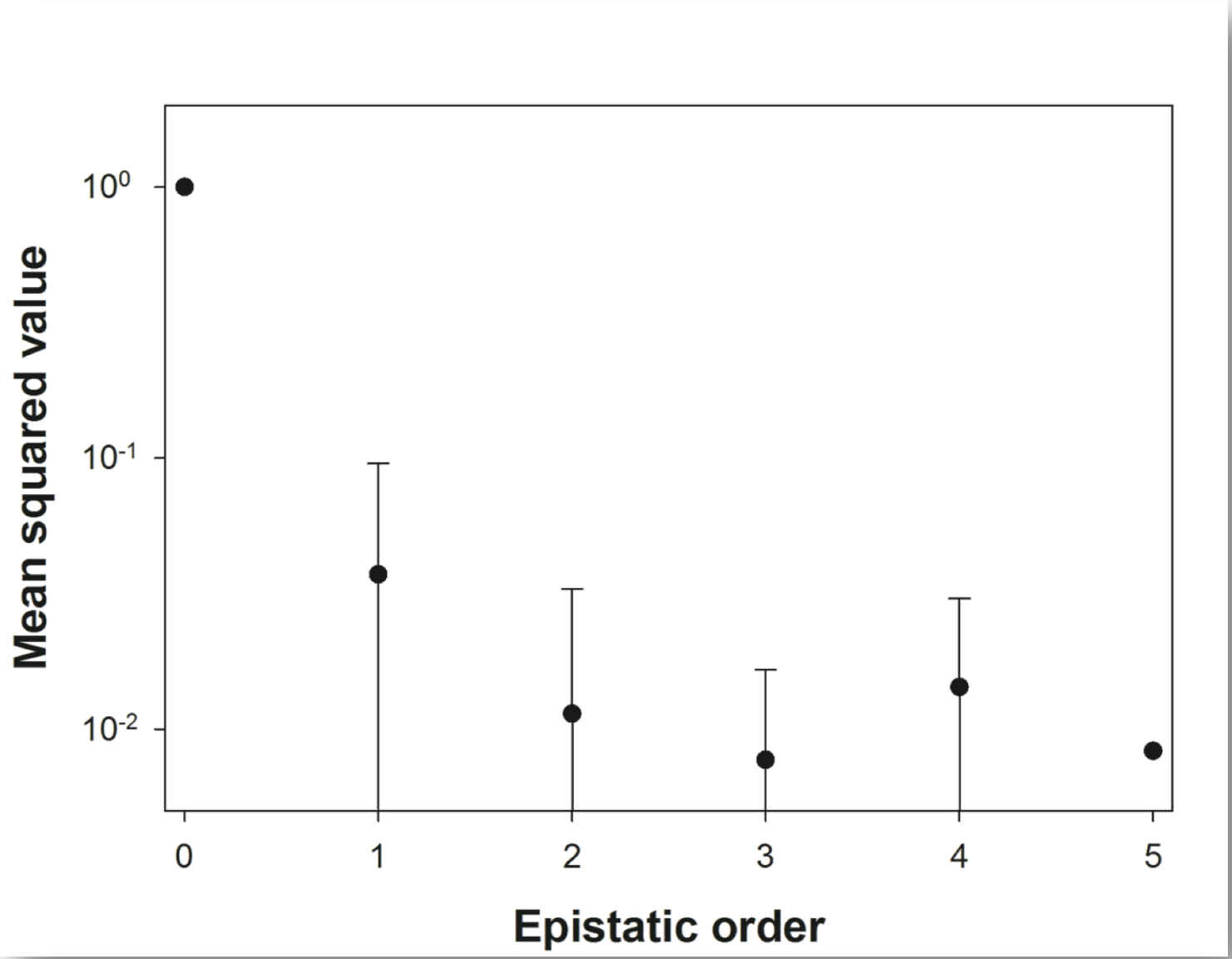
Table 2. Observed fitness and epistasis coefficients (ε) computed for each genotype carrying at least two mutations.

Genotype	Fitness	Epistasis	*P*	Type of epistasis
11000	0	−1.010±0.086	< 0.001	Magnitude
10100	0	0		
10010	0	−1.023±0.089	< 0.001	Magnitude
10001	0	−0.999±0.067	< 0.001	Magnitude
01100	1.019±0.014	1.019±0.012	< 0.001	Magnitude
01010	1.085±0.019***	−0.041±0.131	0.380	
01001	1.120±0.016***	0.021±0.113	0.392	
00110	1.078±0.022**	1.078±0.015	< 0.001	Sign
00101	1.004±0.006	1.004±0.011	< 0.001	Magnitude
00011	1.078±0.022**	−0.036±0.116	0.380	
11100	0	0		Magnitude (10000 on 01100)
11010	0	−1.078±0.128	< 0.001	Magnitude (10000 on 01010)
11001	1.082±0.022**	0.028±0.112	0.386	Magnitude (00001 on 11000, 01000 on 10001)
10110	1.061±0.018**	1.061±0.015	< 0.001	Magnitude (00010 on 10100), Reciprocal sign (00100 on 10010)
10101	0	−0.962±0.001	< 0.001	Magnitude (10000 on 00101)
10011	1.074±0.028*	0.007±0.115	0.398	Magnitude (00001 on 10010, 00010 on 10001)
01110	1.070±0.030*	1.070±0.015	< 0.001	Magnitude (00100 on 01010)
01101	1.055±0.027	1.043±0.014	< 0.001	Magnitude (00100 on 01001)
01011	1.112±0.021***	−0.062±0.155	0.368	
00111	1.049±0.029	1.049±0.015	< 0.001	Magnitude (00100 on 00011)
11110	1.039±0.021	1.039±0.013	< 0.001	Magnitude (00010 on 11100), Reciprocal sign (00100 on 11010)
11101	0	0		Magnitude (10000 on 01101)
11011	0	−1.125±0.150	< 0.001	Magnitude (10000 on 01011), Reciprocal sign (00010 on 11001, 01000 on 10011)
10111	1.043±0.030	1.043±0.015	< 0.001	Magnitude (00010 on 10101, 00100 on 10011)
01111	1.050±0.027	1.050±0.014	< 0.001	Magnitude (00100 on 01011)
11111	1.012±0.010	1.012±0.012	< 0.001	Magnitude (00010 on 11101), Reciprocal sign (00100 on 11011)

Note.- Two samples *t*-test comparing with the wildtype: * $0.05 > P \geq 0.01$, ** $0.01 > P \geq 0.001$, *** $0.001 > P$. Fitness errors represent ±1 SEM. Epistasis errors represent ±1 SD. *P*-values correspond to the *z*-score test of significance of each epistasis coefficient.

✓ Average of positive sign epistasis, 0.417±0.151 (epistasis=0.007) mutation 00100: reciprocal sign epistasis.

Significant contribution of higher-order epistasis (Walsh coefficients)



Lalić & Elena (2015) *J. Evol. Biol.* **28**, 2236

Conclusions Part II

- ✓ Positive epistasis are the norm in the genome of RNA viruses and small ssDNA viruses (reviewed in Elena *et al.* 2010).
- ✓ Pervasive reciprocal sign epistasis indicates that the fitness landscape for RNA viruses shall be very rugged. The ruggedness of adaptive landscapes is critical to predict whether evolving populations may reach the global optima or, by contrast, may get stuck into suboptimal fitness peaks.

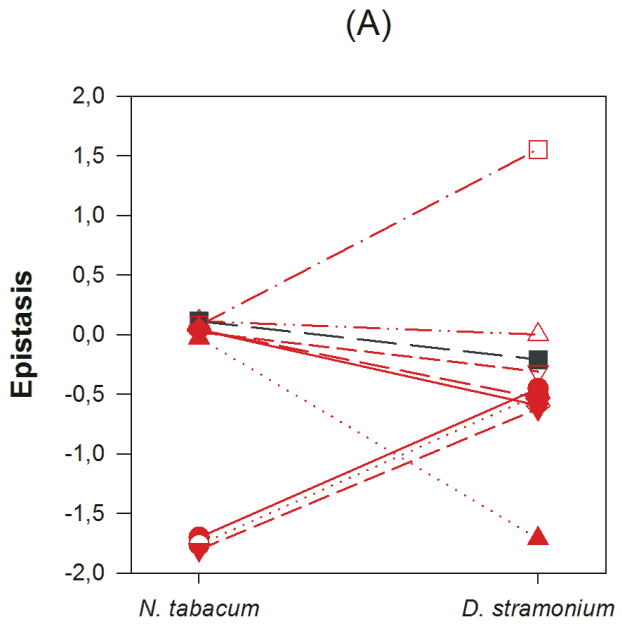
$$W = G + E + G \times G + G \times E$$

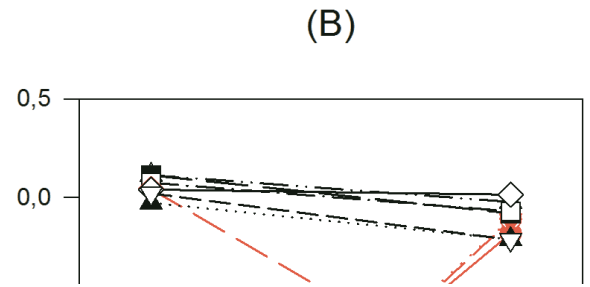
Part III
Testing for pleiotropic epistasis (*G×G×E*)

Table 1. Epistasis of double mutants in each host. Average epistasis was computed after excluding lethal combinations. Sign epistasis refers to cases in which the sign of the fitness effect depends on the genetic background. Reciprocal (recip.) sign epistasis means that the sign of the fitness effect of a mutation is conditional upon the state of another locus and *vice versa*. Last row shows the significance test for the average epistasis. Red numbers indicate significant changes in epistasis from the primary host (*N. tabacum*) to alternative ones (paired *t*-tests corrected for multiple comparisons; figure 1). Errors represent ±1 s.e.m.

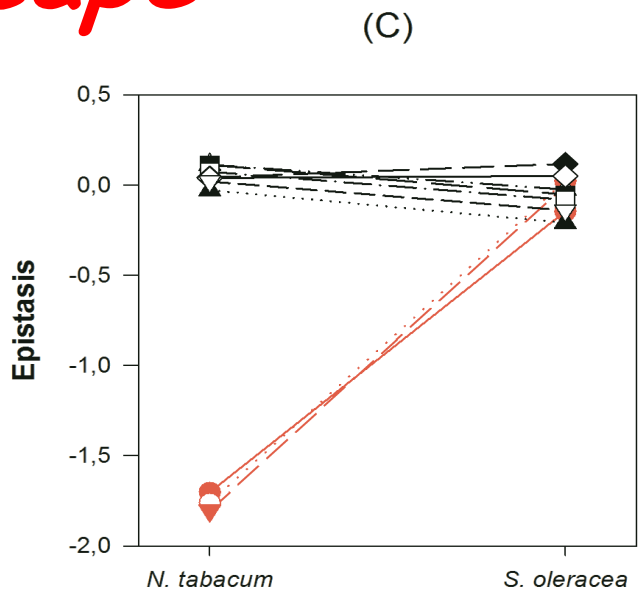
genotype	*N. tabacum*	*D. stramonium*	*H. annuus*	*S. oleracea*	average epistasis
PC6/PC63	0.0730	1.5520[a]	−0.0725	−0.0828	0.3674 ± 0.3965
PC6/PC76	−1.8050[a] (sign)	− 0.6233 (sign)	− 0.1178	− 0.0055	−0.6379 ± 0.4116
PC19/PC41	0.1117[a] (recip. sign)	0	−0.0245	−0.0263	0.0152 ± 0.0327
PC22/PC69	−0.0293	− 1.7129[a]	−0.2147	−0.2106	−0.5419 ± 0.3927
PC22/PC72	0.0179	− 0.3213[a]	−0.2172	−0.1414	−0.1633 ± 0.0698
PC22/PC95	−1.7024[a]	− 0.4537	−0.1855	− 0.1474	−0.6222 ± 0.3665
PC40/PC83	0.1111	−0.2108	−0.0829	−0.0535	−0.0590 ± 0.0662
PC67/PC76	0.0408	− 0.5341[a] (sign)	− 1.0253[a] (recip. sign)	0.1158	−0.3507 ± 0.2677
PC69/PC76	−1.7620[a]	− 0.5057[a] (sign)	− 0.1112	0.0221	−0.5892 ± 0.4067
PC76/PC95	0.0381	− 0.5955[a]	0.0127	0.0496	−0.1238 ± 0.1574
average epistasis	0.0519 ± 0.0193	−0.2834 ± 0.3187	−0.2185 ± 0.1043	−0.0480 ± 0.0316	
t-test (9 d.f.)	0.0358	0.4034	0.0695	0.1630	

✓ Significant differences among genotypes ($F_{9,177} = 168.593$, $P < 0.001$).
Magnitude of the negative ε_{xy} decreases and tends towards zero as the genetic relatedness between the primary host and the alternative one decreases.





The adaptive *seascape*

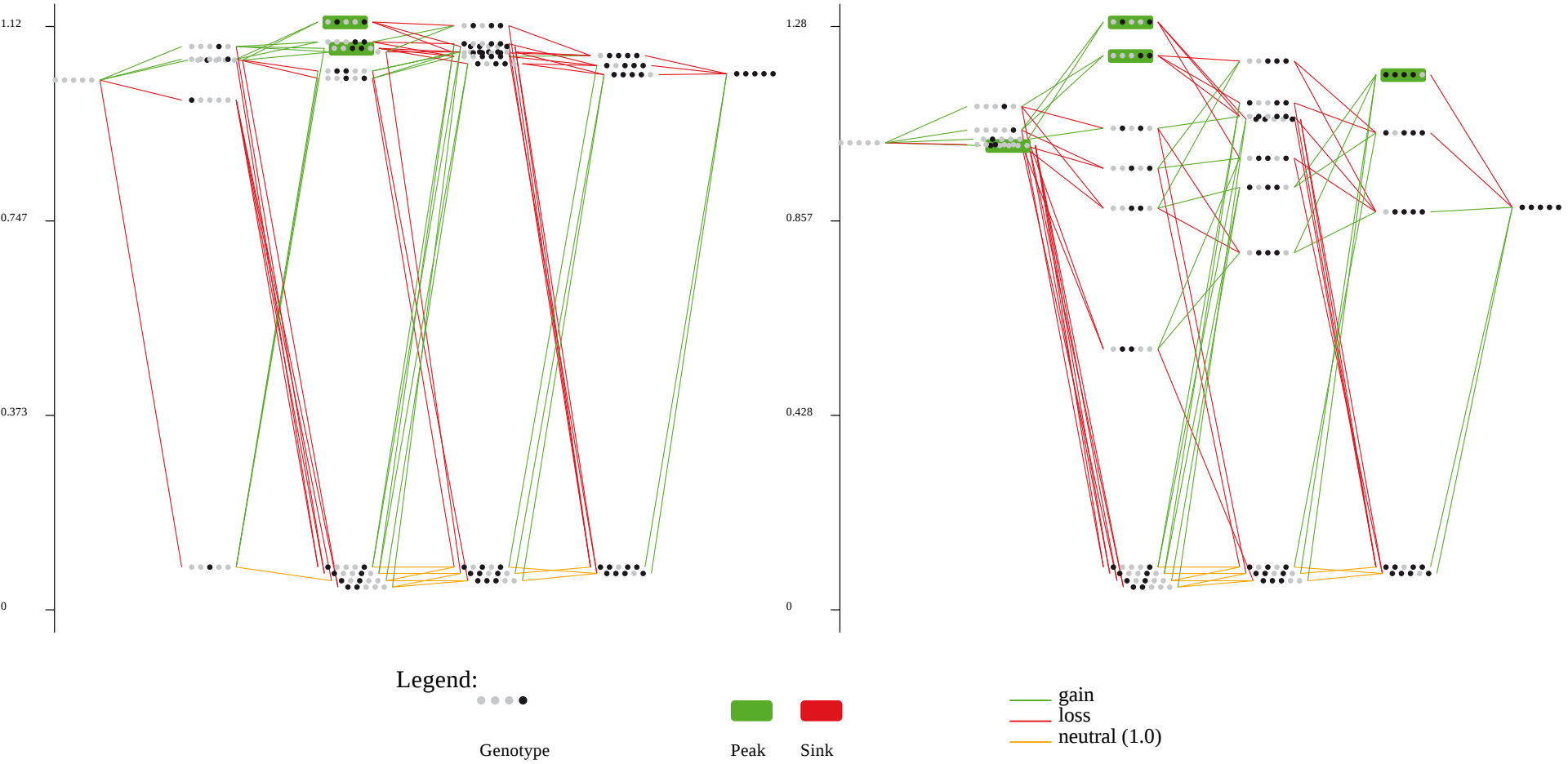


- ✓ Most interestingly, a significant genotype by host effect exists ($F_{27,177}$ = 1.55×10^5, P < 0.001), suggesting that the magnitude of epistasis depends on the host species in which it is measured.

How congruent is the topography in the ancestral and novel hosts?

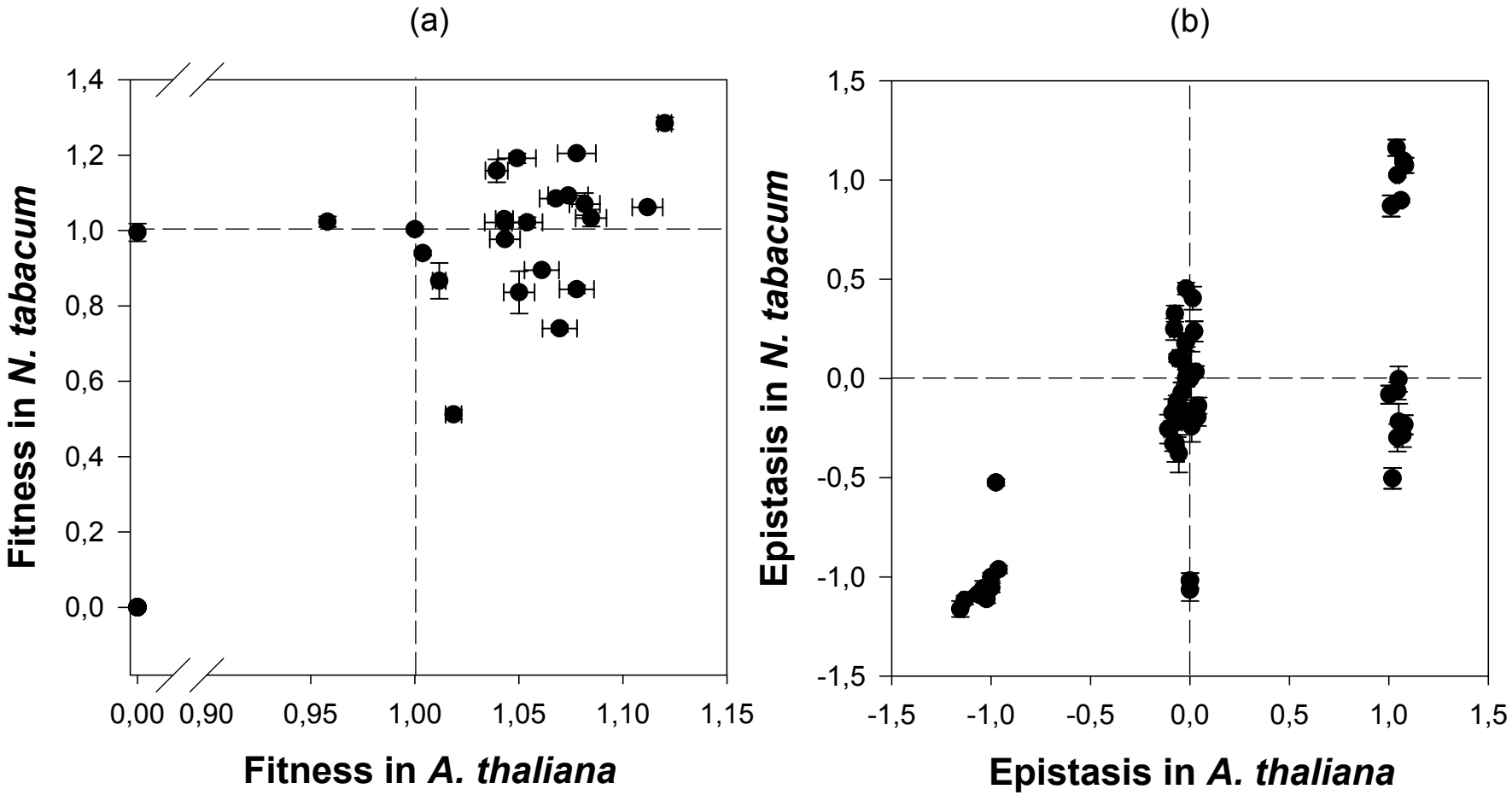
A. thaliana

N. tabacum



Cervera *et al.* (2016) *J. Virol.* **90**, 10160

Pleiotropic fitness effects and variable epistasis



The epistasis transition matrix: the landscape is smoother in the novel host

		A. thaliana				
N. tabacum		No	Magnitude	Sign	Reciprocal sign	
	No	37	8	1	0	46
	Magnitude	2	9	0	0	11
	Sign	1	4	0	2	7
	Reciprocal sign	0	5	3	3	11
		40	26	4	5	

Cervera *et al.* (2016) *J. Virol.* **90**, 10160

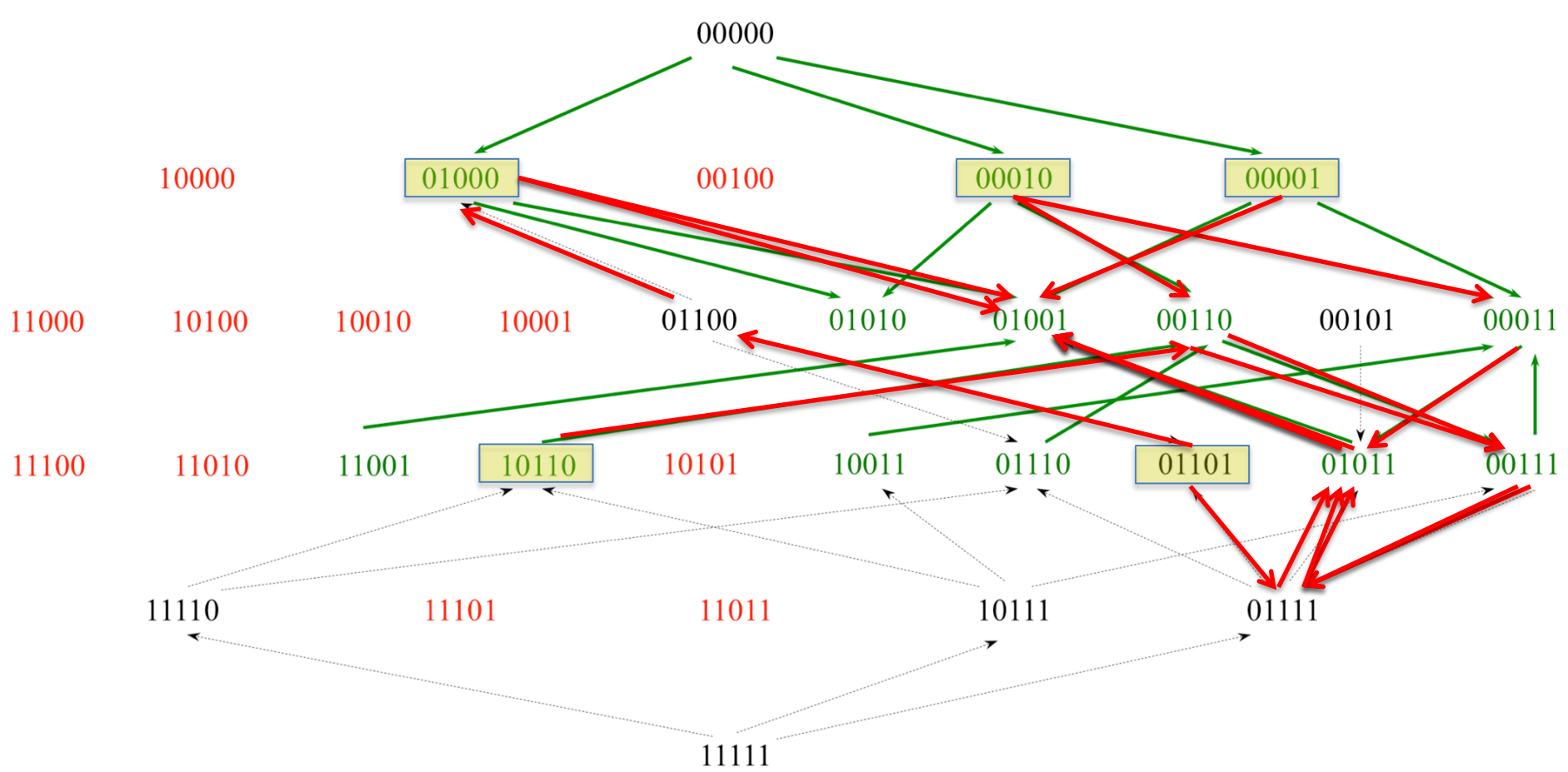
Conclusions Part III

- ✓ The effect of point mutations on TEV fitness is strongly determined by $G \times G \times E$ interactions: the fitness value of a given mutation depends on the genotypic background wherein it appears in as much as on the host infected by the mutant genotype

- ✓ The effect of the host on the strength and sign of epistasis is not stochastic but to a significant extent mediated by the degree of genetic relatedness with the primary host.

$$W = G + E + G \times G + G \times E + G \times G \times E$$

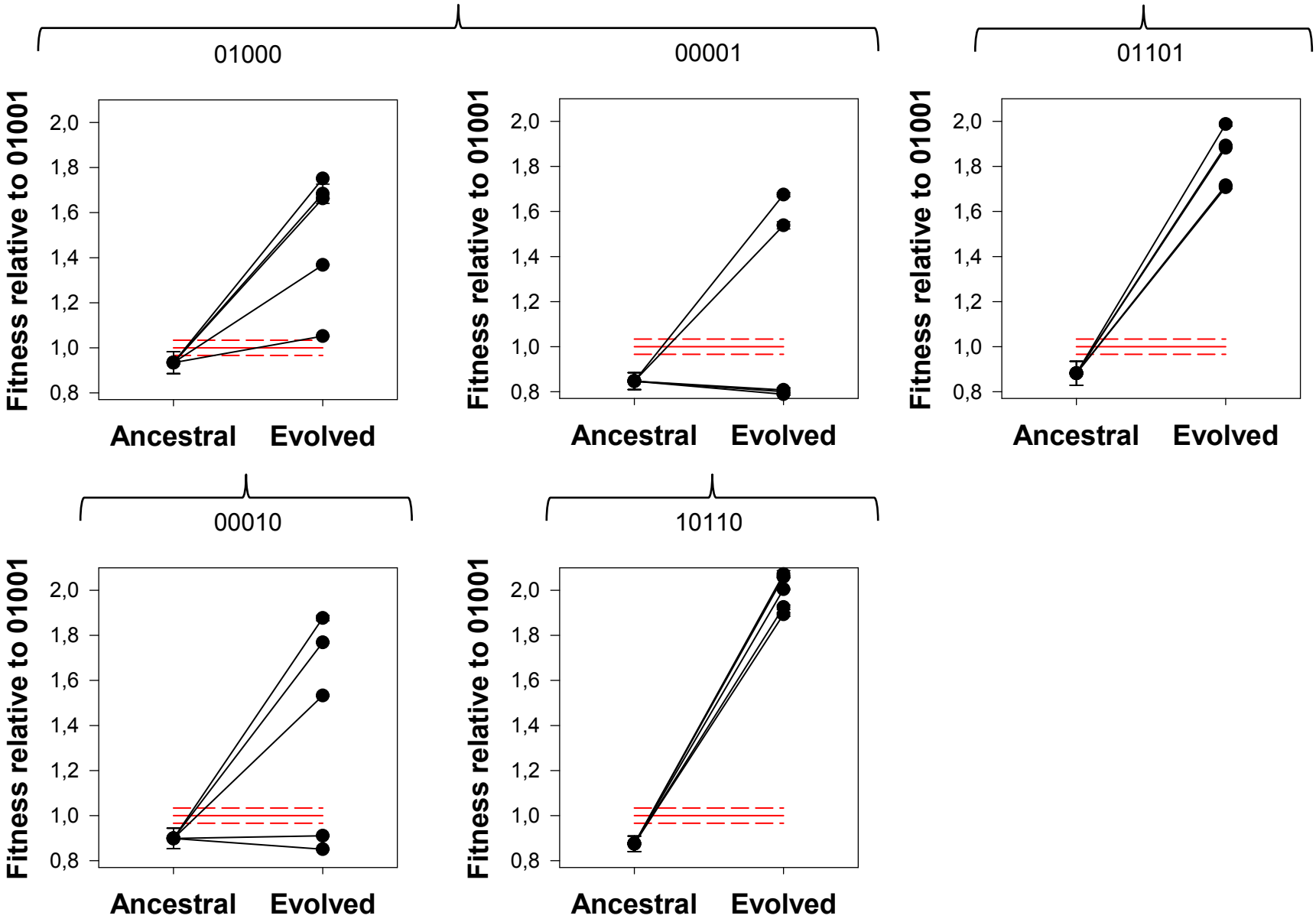
Part IV
Replaying the tape: how predictable/reproducible is virus evolution?



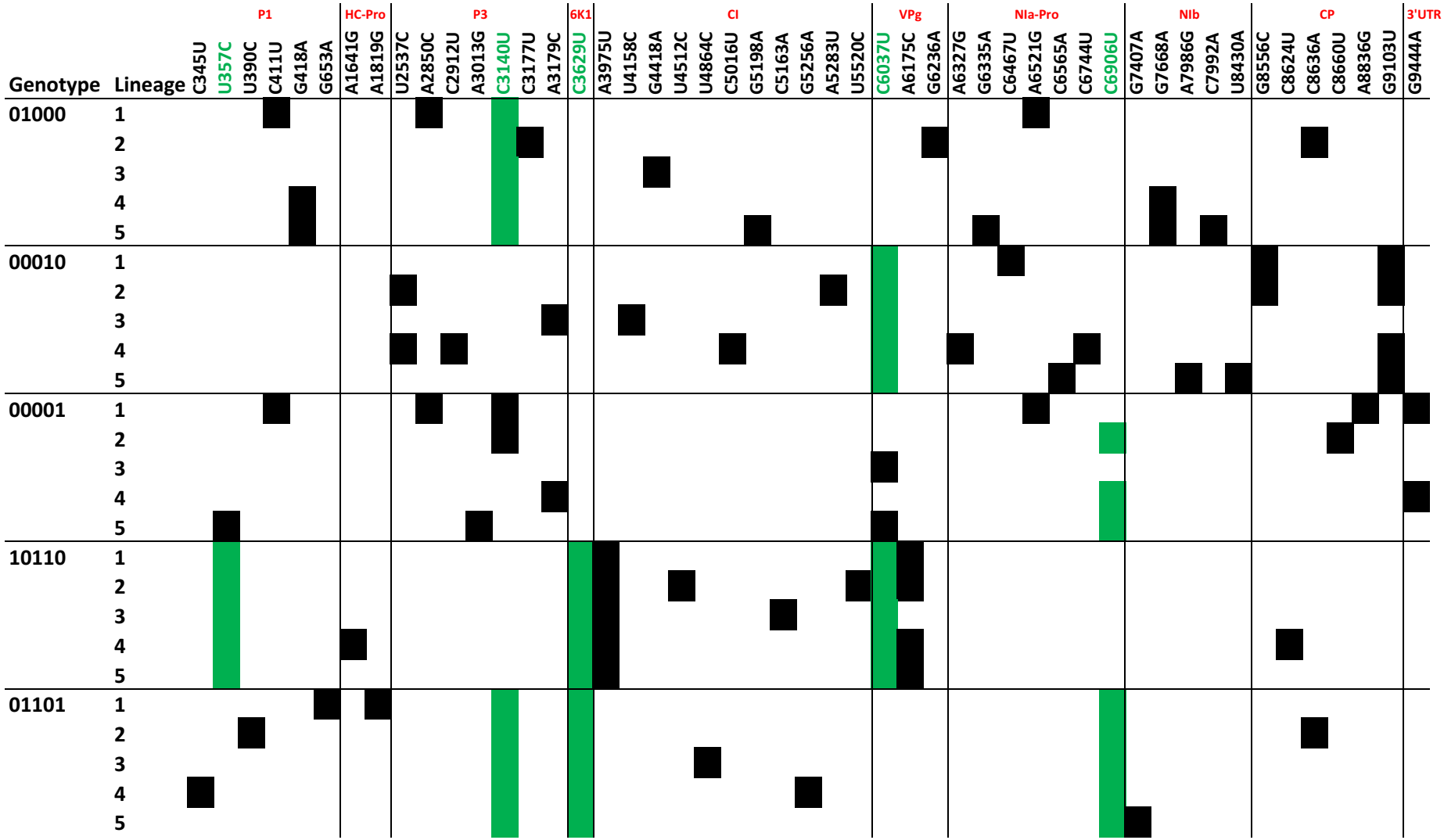
✓ Expected length of the adaptive walk to the optimum genotype:

01000	1 step
00010	2 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 *et al.* (2016) *Proc. R. Soc. B.* **283**, 20160984

How much the outcome of evolution depends on contingency?

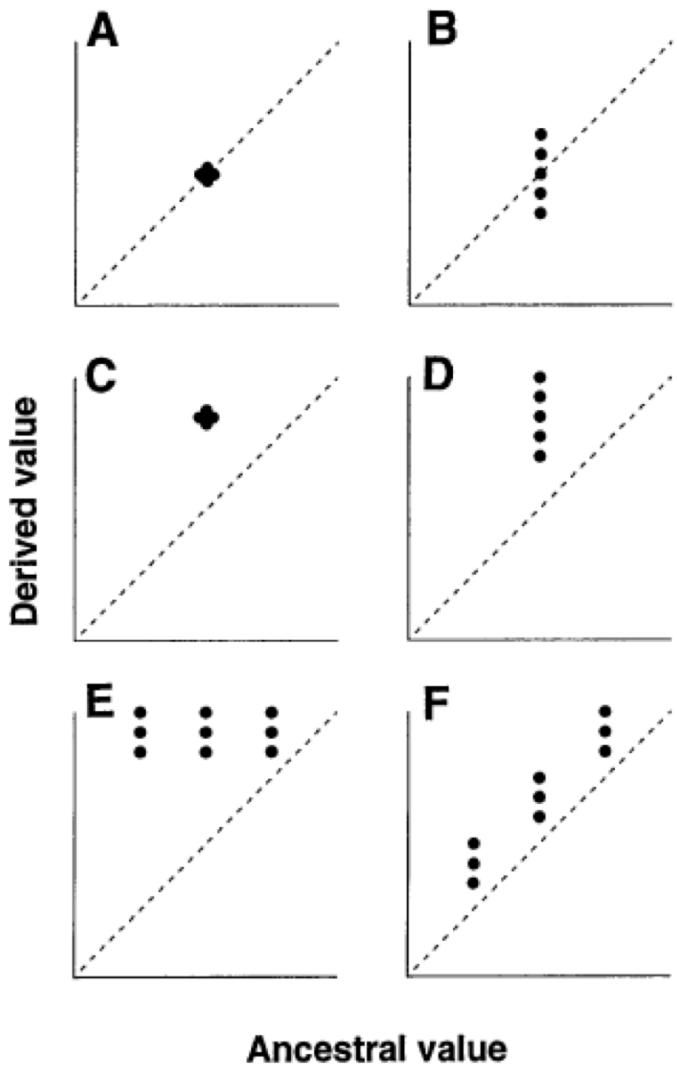
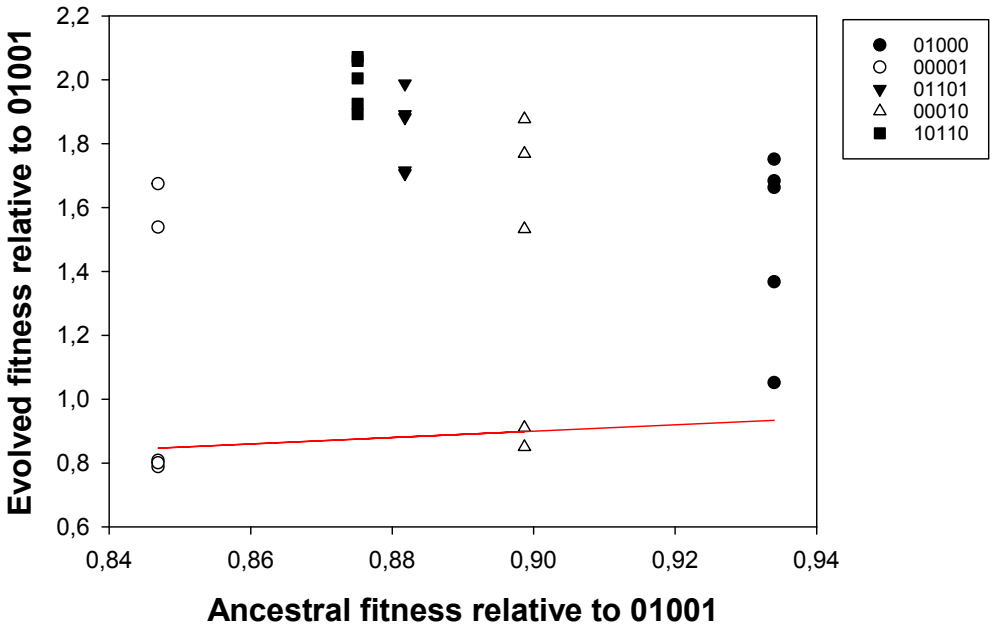
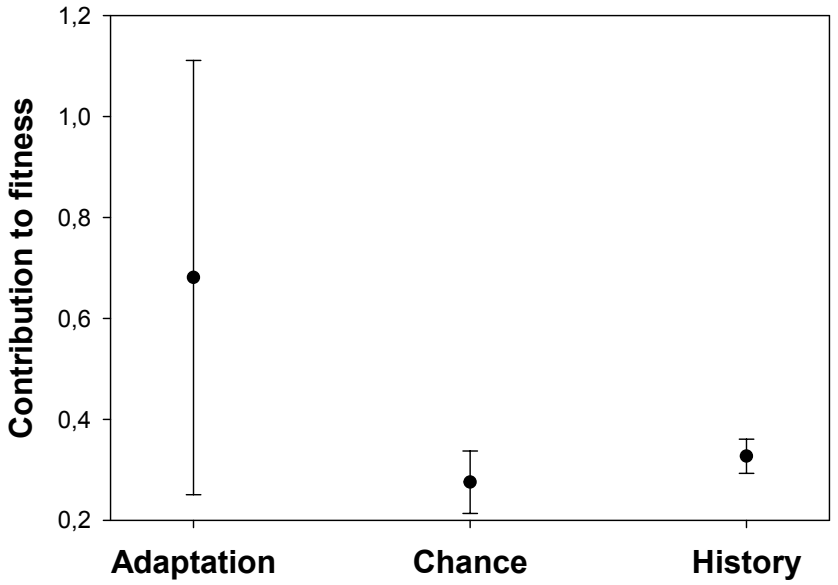


Fig. 1. Schematic representation of effects due to adaptation, chance, and history on evolutionary change and diversification. (**A**) No initial variation and no evolutionary change and hence no effects. (**B**) An effect due to chance only. (**C**) An effect due to adaptation only. (**D**) Effects 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 *et al.* (1995) *Science* **267**, 87



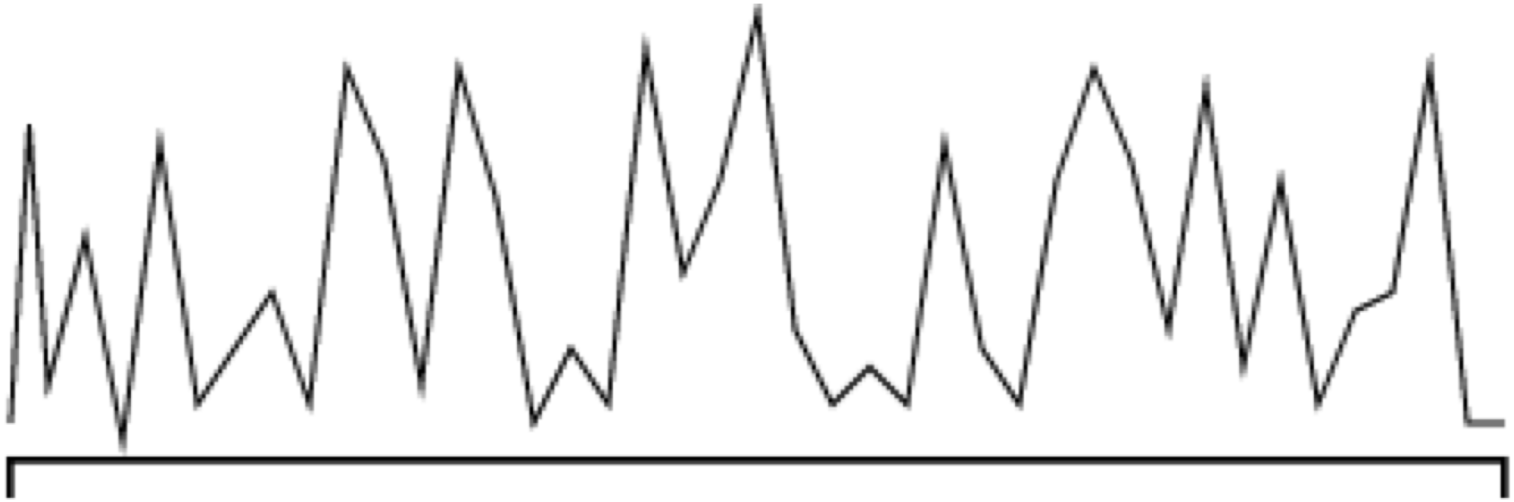


- ✓ Observed slope is different from diagonal (*t*-test, $P = 0.005$): initial differences due to history have been erased by subsequent effects due to chance and adaptation.
- ✓ Relative contributions to adaptation: adaptation > chance ~ history

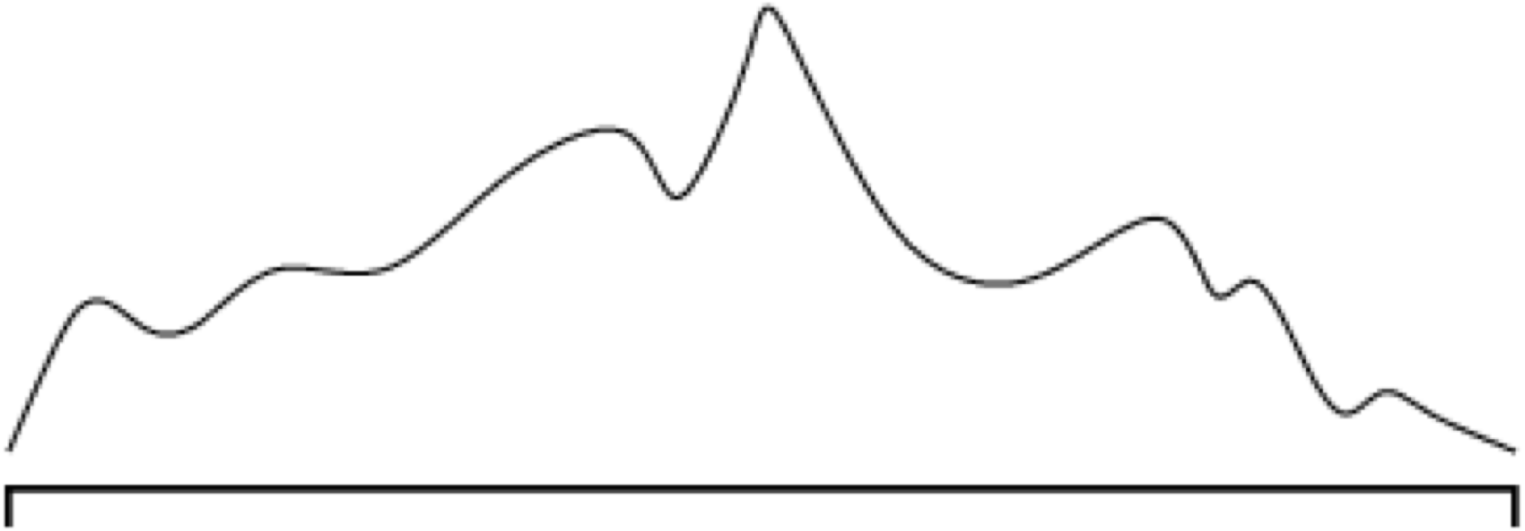
Take home messages

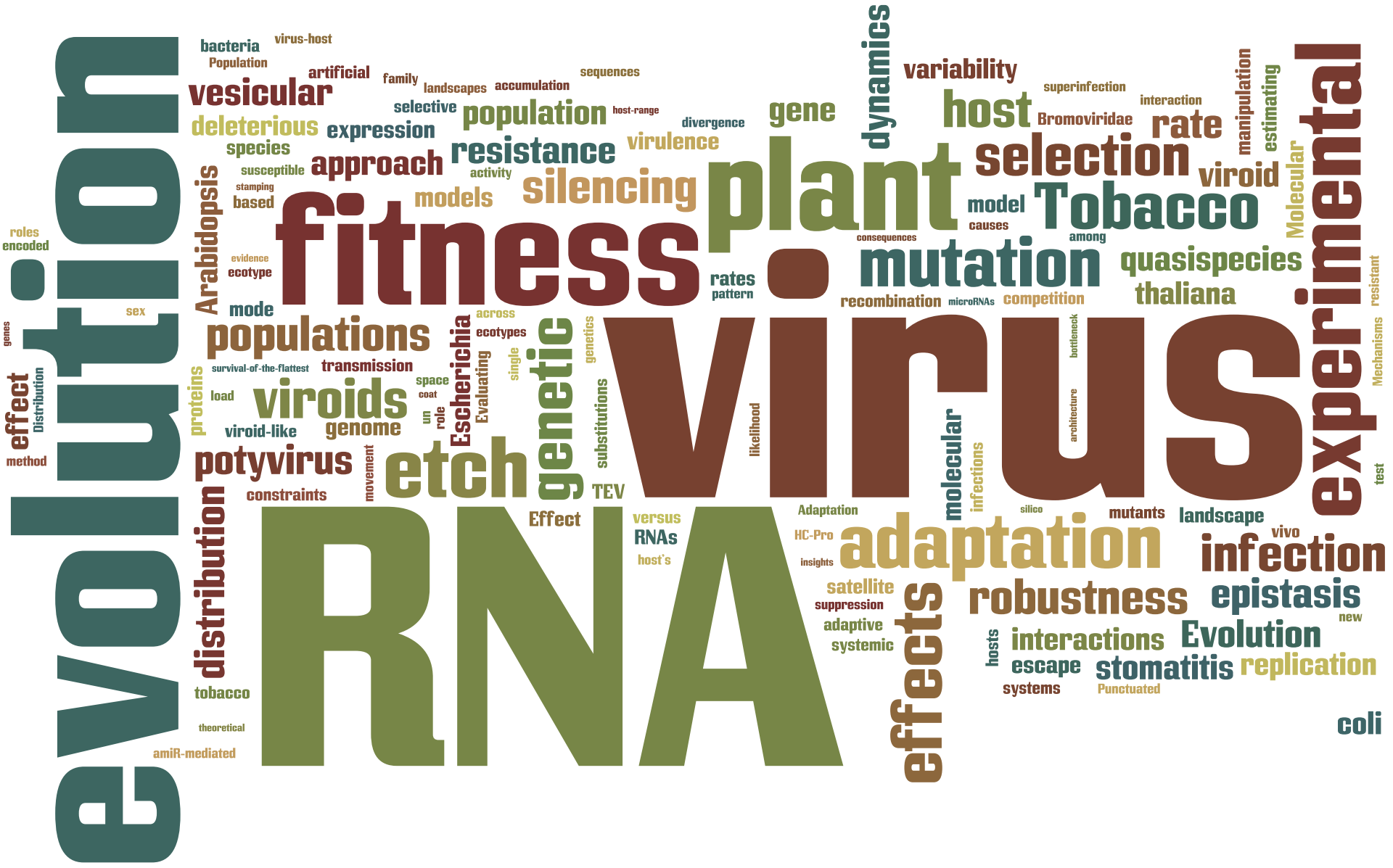
- ✓ Variance in the type and strength of epistasis among pairs of random mutations exist. On average ***positive epistasis is the norm***, as expected for a anti-redundant genome. This has implications in the evolution of robustness and recombination.
- ✓ ***Sign and reciprocal sign epistasis are pervasive*** among random pairs of mutations, depicting a rugged fitness landscape.
- ✓ The magnitude and sign of ***epistasis depends on the host*** being infected.
- ✓ ***High-order epistasis*** significantly contribute to define the topography of fitness landscapes.
- ✓ RNA viruses efficiently explore the fitness landscape and ***escape from local fitness optima***, regardless how far are from their basin of attraction.
- ✓ Escape occurs by fixation of additional mutations at alternative loci.
- ✓ The ***topography of fitness landscapes depends on the host*** being infected: it is smoother in a novel host than in the natural reservoir, allowing access to a better genotype that was not accessible in the reservoir.
- ✓ Both topographies match the expectations from a ***random uncorrelated landscape***: between the House-of-Cards and the rough Mount Fuji models.

House-of-cards



Rough Mt. Fuji




evolution
RNA
virus
fitness
plant
experimental
adaptation
infection
mutation
Tobacco
host
selection
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
genome
transmission
molecular
RNAs
versus
Effect
TEV
constraints
movement
substitutions
likelihood
infections
satellite
suppression
adaptive
systemic
hosts
escape
systems
Punctuated
tobacco
theoretical
amiR-mediated
coli
bacteria
virus-host
Population
artificial
family
landscapes
accumulation
sequences
selective
host-range
divergence
gene
virulence
species
susceptible
stamping
activity
evidence
ecotype
rates
pattern
recombination
microRNAs
competition
consequences
among
model
causes
superinfection
interaction
Bromoviridae
manipulation
estimating
roles
encoded
sex
mode
across
ecotypes
survival-of-the-flattest
space
coat
un
role
Evaluating
single
genetics
bottleneck
architecture
proteins
load
viroid-like
effect
Distribution
method
genes
Mechanisms
resistant
test
Adaptation
HC-Pro
insights
silico
mutants
landscape
vivo
new
coli