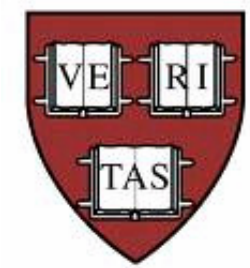
Gene interactions and the geometry of fitness landscapes



Niko Beerenwinkel

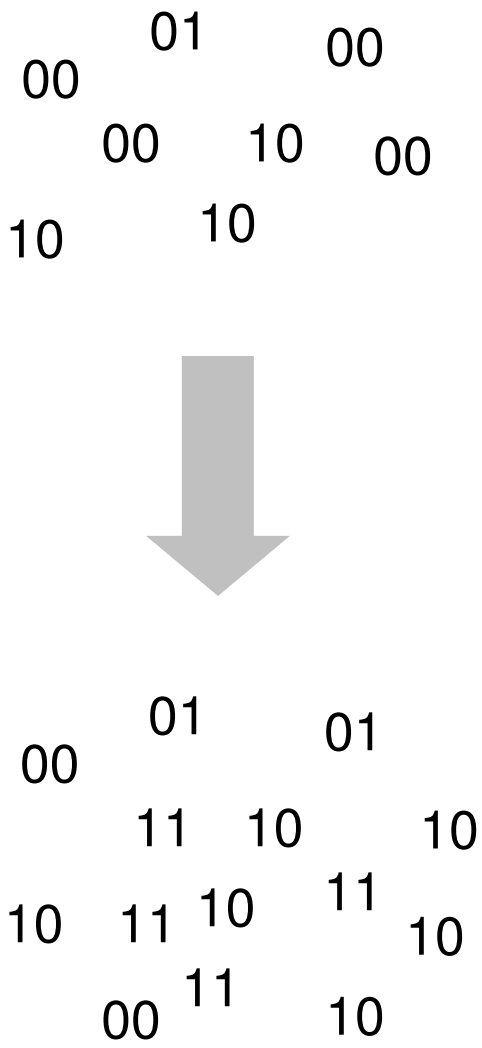
Harvard University

Program for Evolutionary Dynamics

http://www.fas.harvard.edu/~beerenw/

joint work with Lior Pachter, Bernd Sturmfels,
Richard E. Lenski, and Santiago F. Elena

Evolutionary dynamics



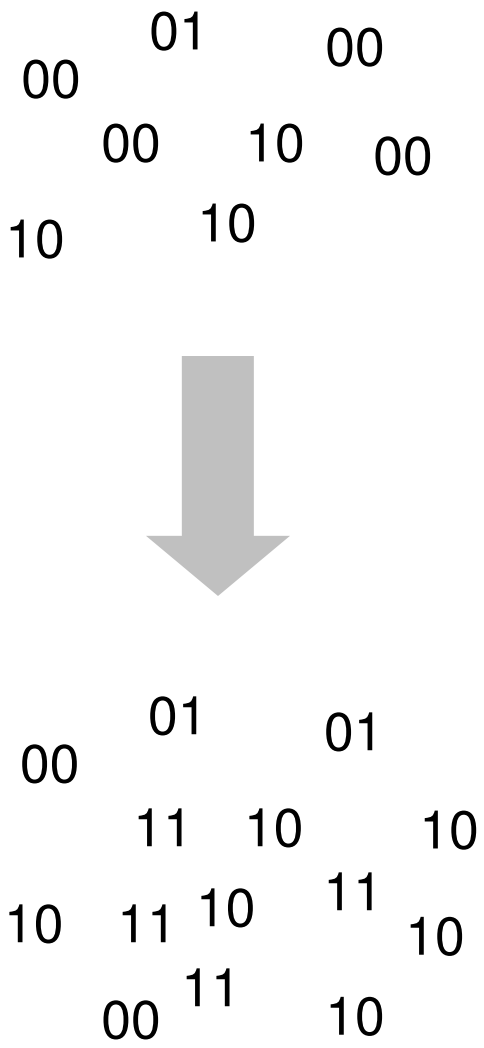
Replication

Mutation

Selection

Recombination

Evolutionary dynamics



Replication

Mutation

Selection

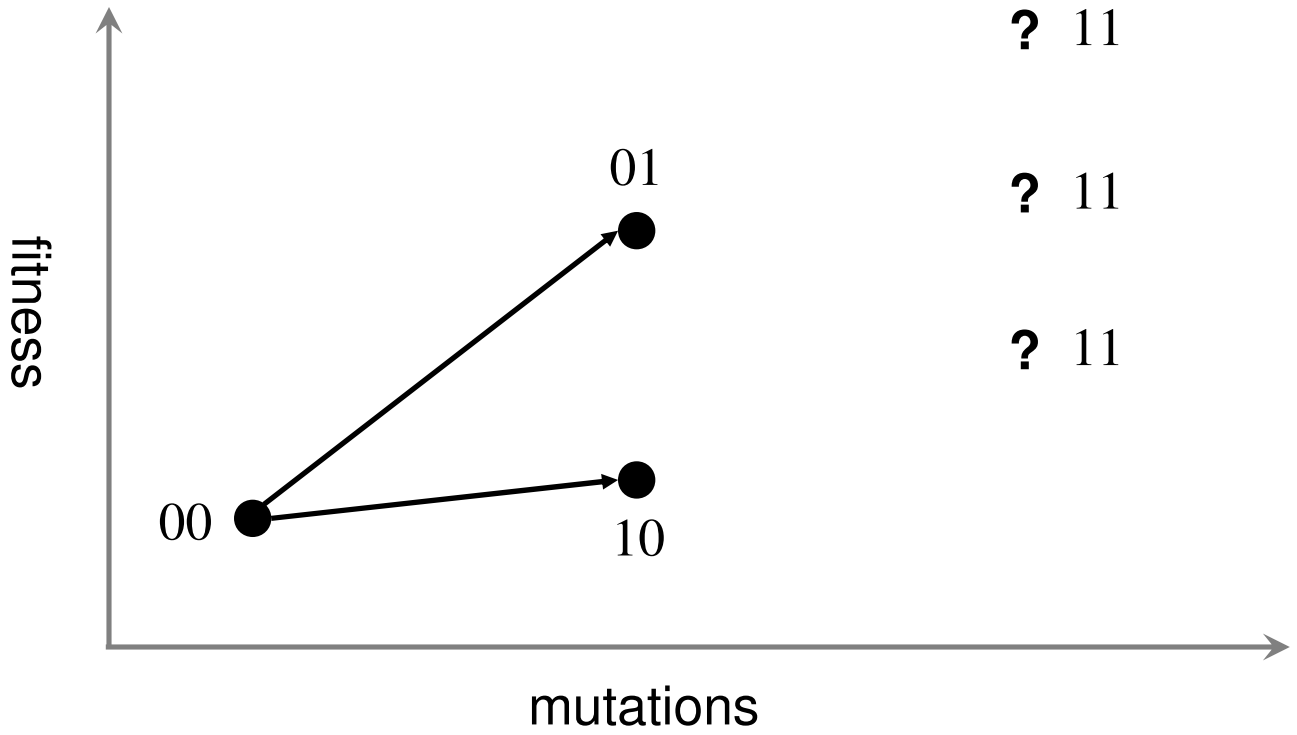
Recombination

$$\Pr(ij \mid t+1) = \Pr(ij \mid t) \cdot w_{ij}$$

w_{ij} = fitness of genotype ij

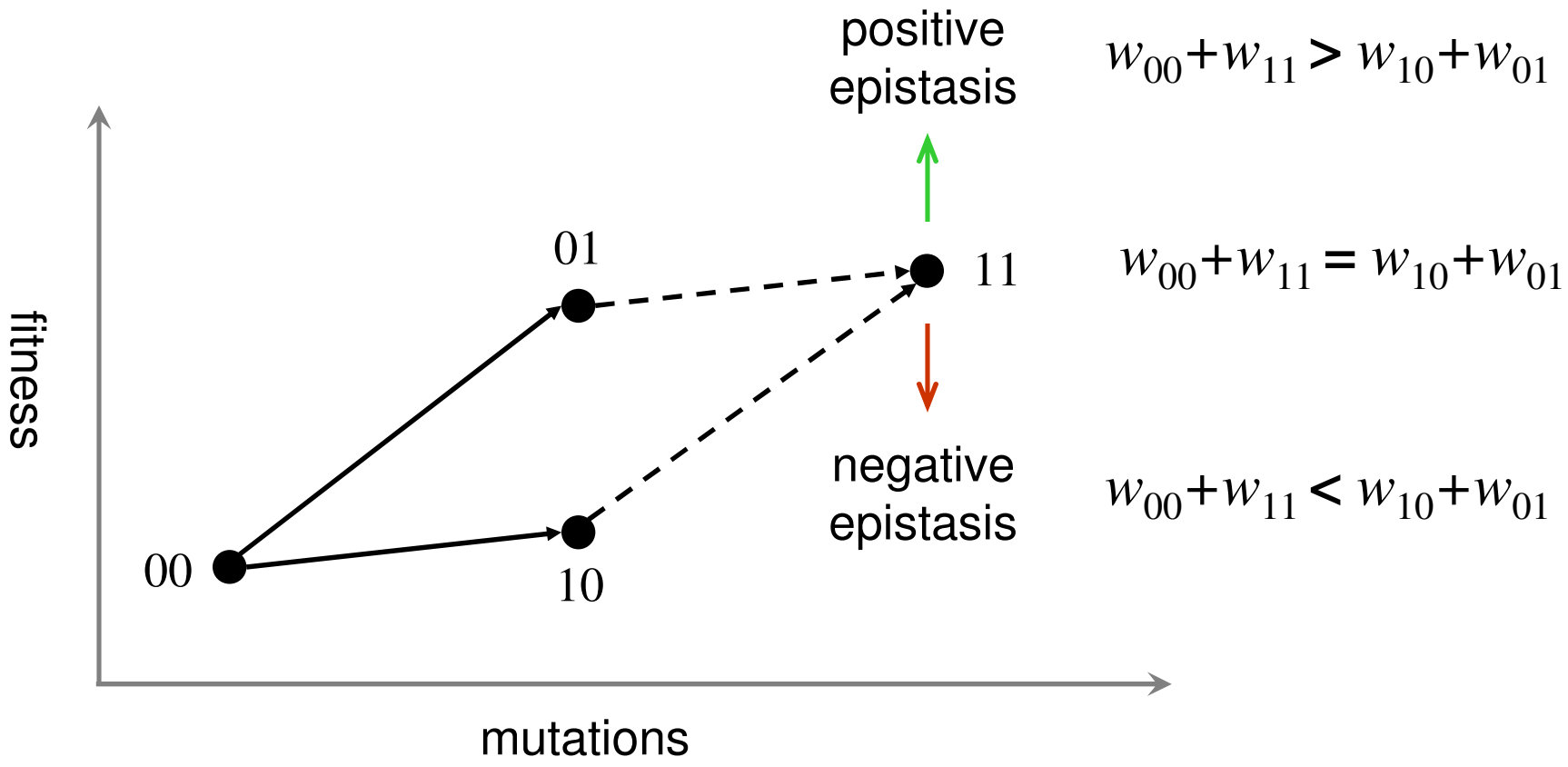
Epistasis

- Two-locus two-alleles: 00 01 10 11
  with fitness landscape w_{00} w_{01} w_{10} w_{11}



Epistasis

- Two-locus two-alleles: 00 01 10 11
  with fitness landscape w_{00} w_{01} w_{10} w_{11}



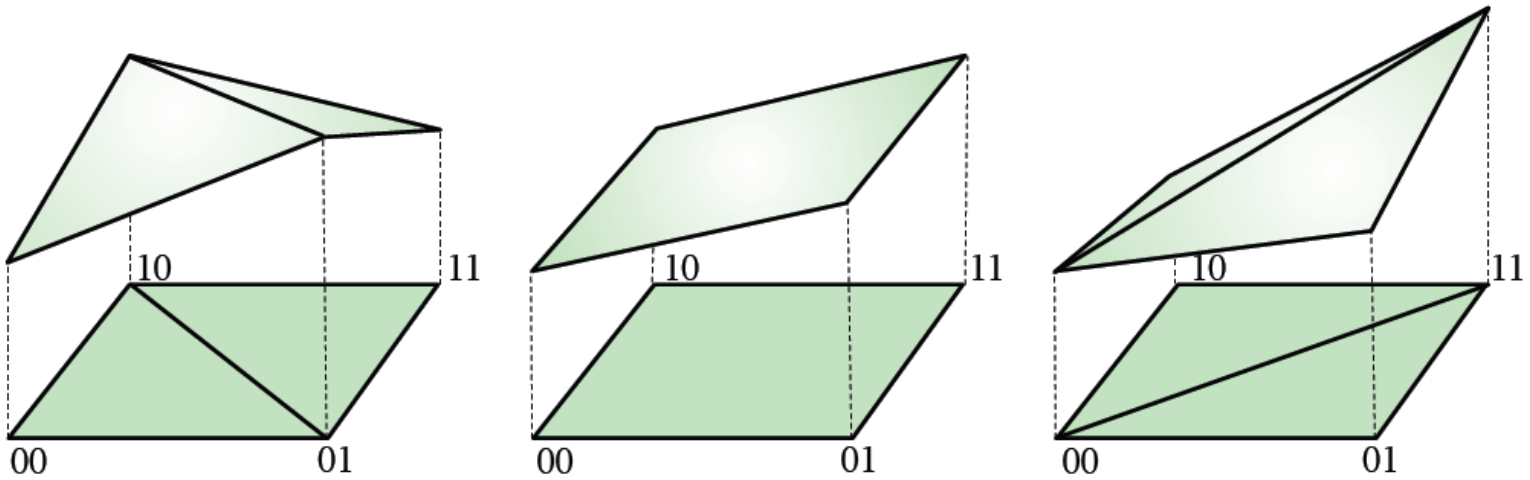
Why does epistasis matter?

- Epistasis is abundant in nature.
- Epistasis can point to interesting gene interactions.
- Epistasis affects the course of evolution, for example:
  - Mutation load
  - Drift and fixation of deleterious mutations
  - Sympatric speciation (vs. allopatric)
  - In some models, recombination is advantageous under negative epistasis, where the advantage can refer to:
    - population fitness at equilibrium
    - time to appearance or fixation of an advantageous type
    - increasing frequency of a modifier allele

Geometric perspective

- Two-locus two-alleles: $00 \quad 01 \quad 10 \quad 11$
  with fitness landscape $w_{00} \quad w_{01} \quad w_{10} \quad w_{11}$

epistasis $u = w_{00} + w_{11} - w_{01} - w_{10}$

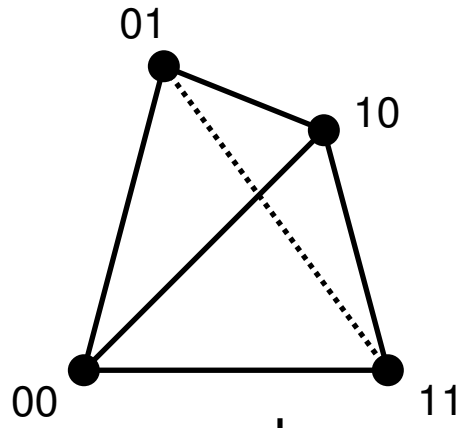


$u < 0$ $\qquad$ $u = 0$ $\qquad$ $u > 0$

Three shapes of fitness landscapes!

The genotope

$$\mathcal{G} = \{00, 01, 10, 11\} \subset$$

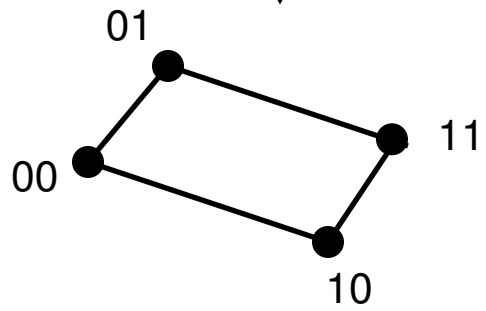


population simplex

ρ

marginalization map

$$\Pi_{\mathcal{G}} =$$



genotope

allele frequency space

Fitness landscapes and gene interactions

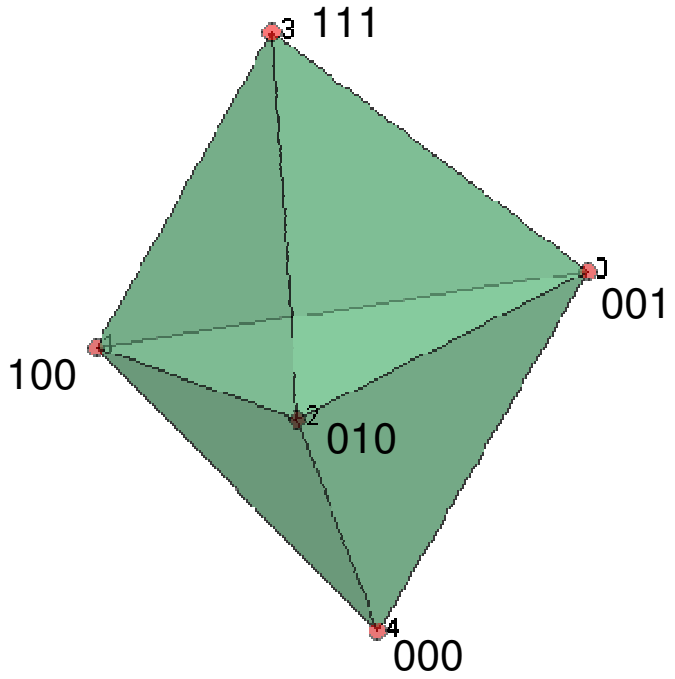
- A fitness landscape is a function $w\colon \mathcal{G} \to \mathbb{R}$.
- Linear functions have no interactions, so consider the interaction space

$$\mathcal{I}_{\mathcal{G}} \quad := \quad (\mathbb{R}^{\mathcal{G}} / \mathcal{L}_{\mathcal{G}})^*$$

- For example: $\mathcal{I}_{\{00,01,10,11\}} = \langle w_{00} + w_{11} - w_{01} - w_{10} \rangle$
- Hypercubes have natural interaction coordinates given by the discrete Fourier transform.
- The interaction space is spanned redundantly by the circuits, i.e., the linear forms with minimal support in $\mathbb{R}^{\mathcal{G}}$.

Example: $\mathcal{G} = \{000, 001, 010, 100, 111\}$

$$\Pi_{\mathcal{G}} =$$

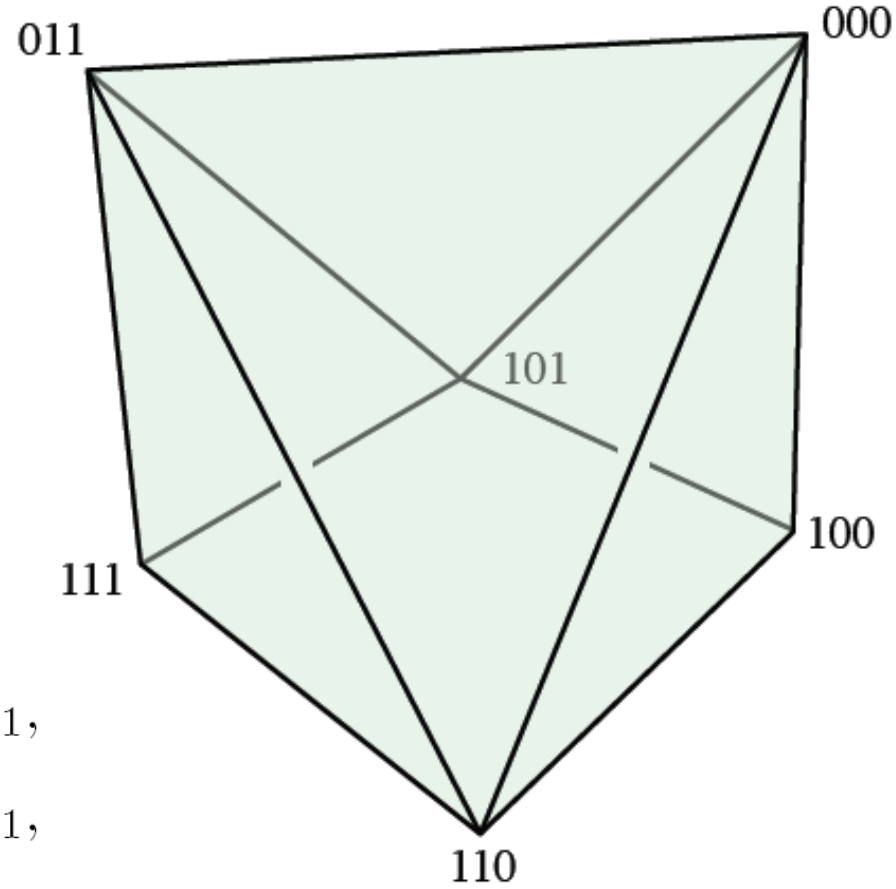


<u>One circuit:</u>

$$f = w_{001} + w_{010} + w_{100} - w_{111} - 2w_{000}$$

Example: $\mathcal{G} = \{000, 110, 011, 100, 101, 111\}$

$$\Pi_{\mathcal{G}} =$$



Four circuits:

$$\begin{aligned} f &= w_{100} - w_{101} - w_{110} + w_{111}, \\ g &= w_{000} - w_{011} - w_{100} + w_{111}, \\ n &= w_{011} + w_{101} + w_{110} - w_{000} - 2w_{111}, \\ s &= w_{000} + w_{101} + w_{110} - w_{011} - 2w_{100}. \end{aligned}$$

The shape of a fitness landscape

- Extend $w\colon \mathcal{G} \to \mathbb{R}$ to the genotope: For all $v \in \Pi_{\mathcal{G}}$,

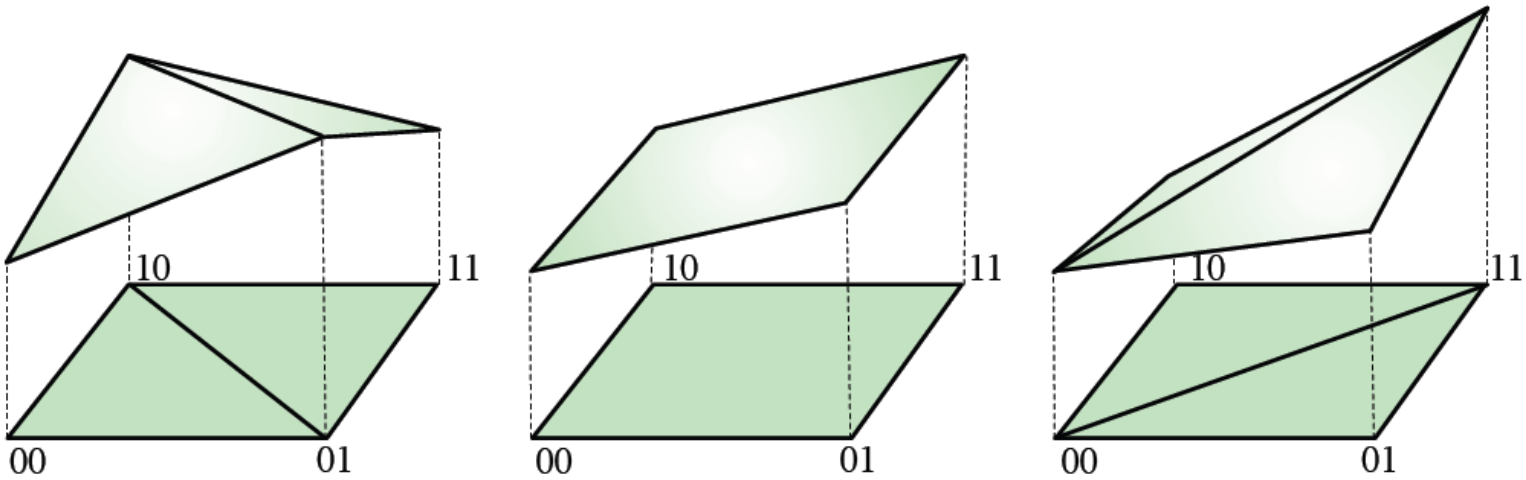
$$\tilde{w}(v) \quad := \quad \max\left\{p \cdot w \ : \ p \in \rho^{-1}(v)\right\}$$

population
fitness

- The continuous landscape is convex and piecewise linear.
- The domains of linearity are the cells in a regular polyhedral subdivision of the genotope.
- Def: This subdivision is the shape of the fitness landscape.
- The circuit sign pattern determines the shape.

Example: $\mathcal{G} = \{00, 01, 10, 11\}$

$$u \quad = \quad w_{00} + w_{11} - w_{01} - w_{10}$$



u < 0 ***u = 0*** ***u > 0***

Fittest populations with fixed allele frequencies:

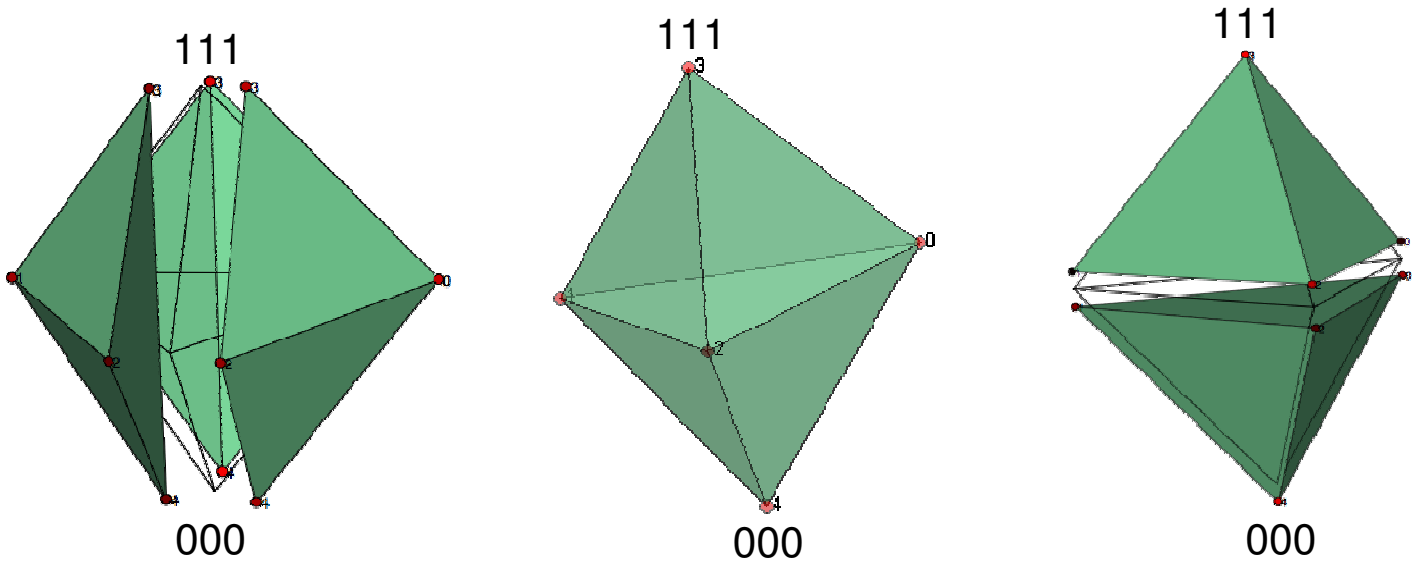
{00, 01, 10}
{01, 10, 11}

{00, 01, 10, 11}

{00, 01, 11}
{00, 10, 11}

Example: $\mathcal{G} = \{000, 001, 010, 100, 111\}$

$$f = w_{001} + w_{010} + w_{100} - w_{111} - 2w_{000}$$



$f < 0$ $f = 0$ $f > 0$

Fittest populations with fixed allele frequencies:

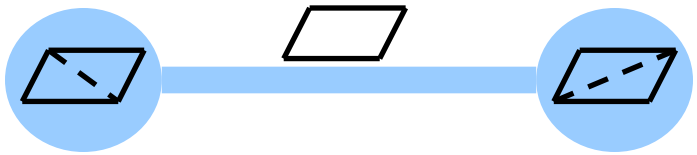
{000, 010, 100, 111}
{000, 001, 100, 111}
{000, 001, 010, 111}

{000, 001, 010, 001, 111}

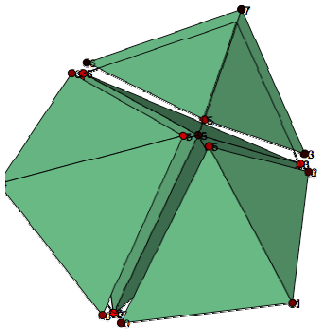
{000, 001, 010, 001}
{001, 010, 001, 111}

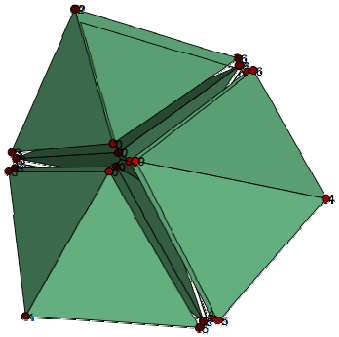
The secondary polytope

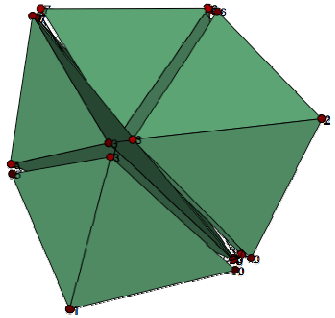
- For a given genotype space, what fitness shapes are there?
- The answer to this *parametric fitness shape problem* is encoded in the secondary polytope.
- For example, the 2-cube has 2 triangulations:



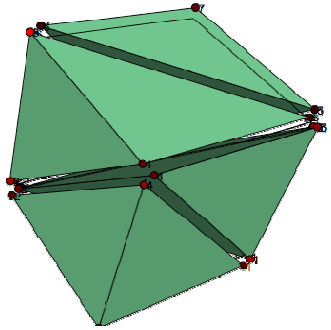
The 74 shapes of the biallelic 3-locus system

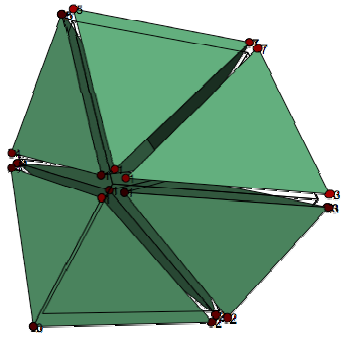


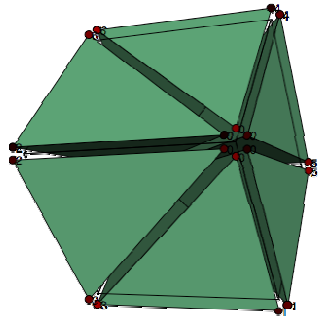




#/T	GKZ	Out-edges	#/T	GKZ	Out-edges
1/1	15515115	3t4q5o6m	38/4	31355313	39$\bar{l}$44$\bar{g}$51c59d
2/1	51151551	7s8r9p10n	39/4	31533513	38l44$\bar{i}$53e60f
3/2	14436114	1$\bar{t}$11b13d17$\bar{e}$	40/4	33155133	42$\bar{j}$45$\bar{g}$54a61b
4/2	14614314	1$\bar{q}$12b14f18$\bar{c}$	41/4	33511533	43$\bar{h}$46$\bar{i}$55a62b
5/2	16414134	1$\bar{o}$15d16f19$\bar{a}$	42/4	35133153	40j45$\bar{k}$57e63f
6/2	34414116	1$\bar{m}$28$\bar{e}$29$\bar{c}$31$\bar{a}$	43/4	35311353	41h46$\bar{k}$58c64d
7/2	41163441	2$\bar{s}$20a22c26$\bar{f}$	44/4	51333315	38g39i65$\bar{b}$68$\bar{a}$
8/2	41341641	2$\bar{r}$21a23e27$\bar{d}$	45/4	53133135	40g42k66$\bar{d}$69$\bar{c}$
9/2	43141461	2$\bar{p}$24c25e30$\bar{b}$	46/4	53311335	41i43k67$\bar{f}$70$\bar{e}$
10/2	61141443	2$\bar{n}$32$\bar{f}$33$\bar{d}$34$\bar{b}$	47/5	13356222	11$\bar{d}$13$\bar{b}$35f71$\bar{e}$
11/3	13446213	3$\bar{b}$12$\bar{l}$47d51$\bar{e}$	48/5	13623522	12$\bar{f}$14$\bar{b}$36d72$\bar{c}$
12/3	13624413	4$\bar{b}$11l48f53$\bar{c}$	49/5	16323252	15$\bar{f}$16$\bar{d}$37b73$\bar{a}$
13/3	14346123	3$\bar{d}$15$\bar{j}$47b54$\bar{e}$	50/5	22265331	20$\bar{c}$22$\bar{a}$35e71$\bar{f}$
14/3	14613423	4$\bar{f}$16$\bar{h}$48b55$\bar{c}$	51/5	22356213	11e17$\bar{b}$38$\bar{c}$71d
15/3	16324143	5$\bar{d}$13j49f57$\bar{a}$	52/5	22532631	21$\bar{e}$23$\bar{a}$36c72$\bar{d}$
16/3	16413243	5$\bar{f}$14h49d58$\bar{a}$	53/5	22623513	12c18$\bar{b}$39$\bar{e}$72f
17/3	23346114	3e28$\bar{g}$51b54d	54/5	23256123	13e17$\bar{d}$40$\bar{a}$71b
18/3	23613414	4c29$\bar{i}$53b55f	55/5	23612523	14c18$\bar{f}$41$\bar{a}$72b
19/3	26313144	5a31$\bar{k}$57d58f	56/5	25232361	24$\bar{e}$25$\bar{c}$37a73$\bar{b}$
20/3	31264431	7$\bar{a}$21$\bar{l}$50c59$\bar{f}$	57/5	26223153	15a19$\bar{d}$43$\bar{e}$73f
21/3	31442631	8$\bar{a}$20l52e60$\bar{d}$	58/5	26312253	16a19$\bar{f}$43$\bar{c}$73d
22/3	32164341	7$\bar{c}$24$\bar{j}$50a61$\bar{f}$	59/5	31265322	20f26$\bar{a}$38$\bar{d}$71c
23/3	32431641	8$\bar{e}$25$\bar{h}$52a62$\bar{d}$	60/5	31532622	21d27$\bar{a}$39$\bar{f}$72e
24/3	34142361	9$\bar{c}$22j56e63$\bar{b}$	61/5	32165232	22f26$\bar{c}$40$\bar{b}$71a
25/3	34231461	9$\bar{e}$23h56c64$\bar{b}$	62/5	32521632	23d27$\bar{e}$41$\bar{b}$72a
26/3	41164332	7f32$\bar{g}$59a61c	63/5	35132262	24b30$\bar{c}$42$\bar{f}$73e
27/3	41431632	8d33$\bar{i}$60a62e	64/5	35221362	25b30$\bar{e}$32$\bar{d}$73c
28/3	43324116	6e17g65$\bar{c}$66$\bar{a}$	65/5	52323216	28c29e44b74$\bar{a}$
29/3	43413216	6c18i65$\bar{e}$67$\bar{a}$	66/5	53223126	28a31e45d74$\bar{c}$
30/3	44131362	9b34$\bar{k}$63c64e	67/5	53312226	29a31c46f74$\bar{e}$
31/3	44313126	6a19k66$\bar{e}$67$\bar{c}$	68/5	61232325	32d33f44a74$\bar{b}$
32/3	61142334	10f26g68$\bar{d}$69$\bar{b}$	69/5	62132235	32b34f45c74$\bar{d}$
33/3	61231434	10d27i68$\bar{f}$70$\bar{b}$	70/5	62221335	33b34d46e74$\bar{f}$
34/3	62131344	10b30k69$\bar{f}$70$\bar{d}$	71/6	22266222	47e50f51$\bar{d}$54$\bar{b}$59$\bar{c}$61$\bar{a}$
35/4	13355331	36$\bar{l}$37$\bar{j}$47$\bar{f}$50$\bar{e}$	72/6	22622622	48c52d53$\bar{f}$55$\bar{b}$60$\bar{e}$62$\bar{a}$
36/4	13533531	35l37$\bar{h}$48$\bar{d}$52$\bar{c}$	73/6	26222262	49a56b57$\bar{f}$58$\bar{d}$63$\bar{e}$64$\bar{c}$
37/4	15333351	35j36h49$\bar{b}$56$\bar{a}$	74/6	62222226	65a66c67e68b69d70f







A biallelic three-locus system in HIV

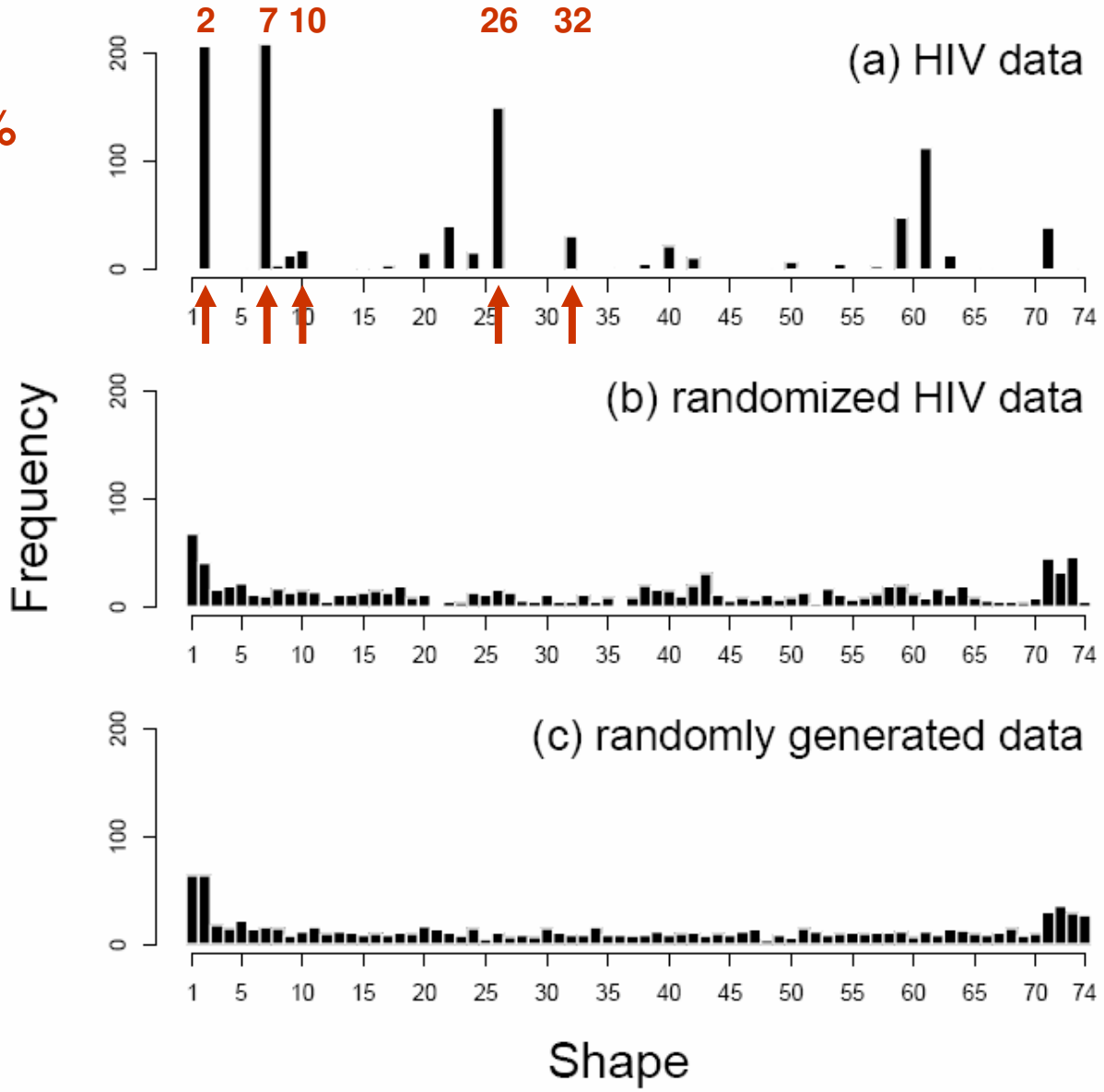
- HIV protease: L90M; RT: M184V and T215Y.
- Fitness measured in single replication cycle, 288 data points (Segal et al., 2004; Bonhoeffer et al., 2004).

- Conditional epistasis:

Circuit	Pair	Context	Cond. epist.	P-value
a	90–184	T215	0.300	0.110
b	90–184	215Y	–0.421	0.059
c	90–215	M184	0.175	0.230
d	90–215	184V	–0.545	0.013
e	184–215	L90	0.682	0.008
f	184–215	90M	–0.039	0.410

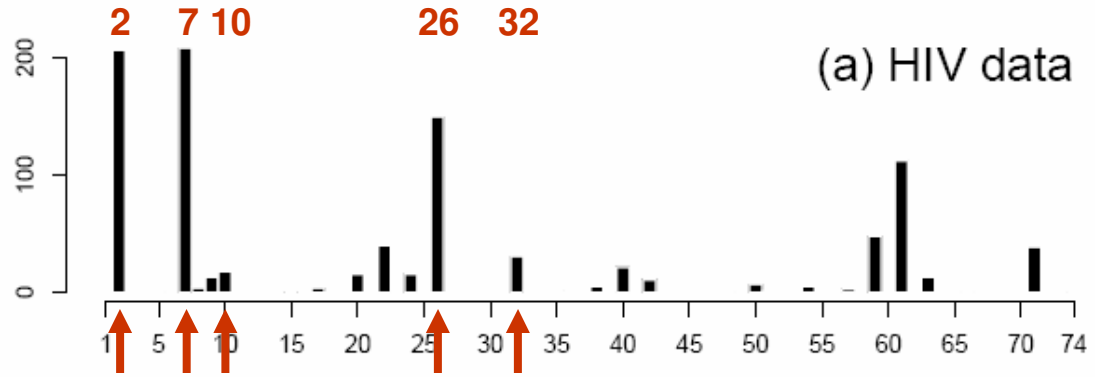
HIV random fitness landscape

> 60%



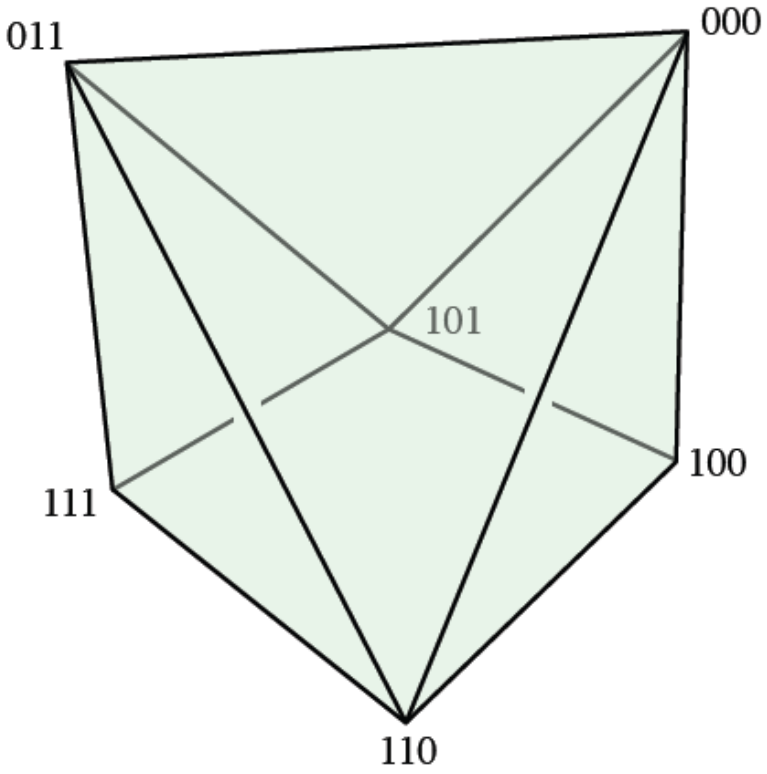
HIV random fitness landscape

> 60%



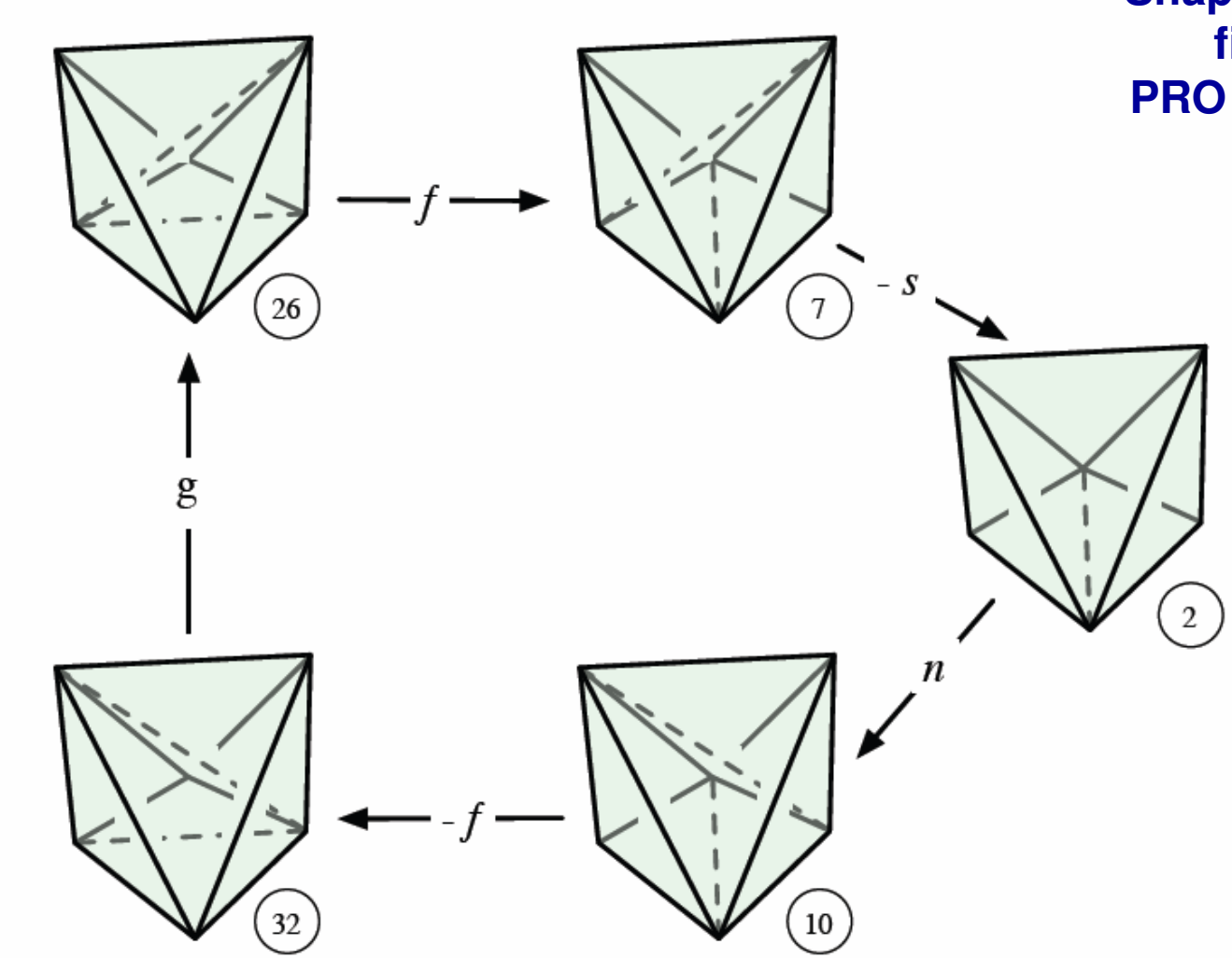
In these five shapes, both 001 and 010 are “sliced off” by the triangulations, i.e., the fittest populations avoid the single mutants {M184V} and {T215Y}.

Hence we consider
000, 011, 100, 101, 110, 111:



HIV secondary polytope

Shape of the HIV random fitness landscape on PRO 90 / RT 184 / RT 215

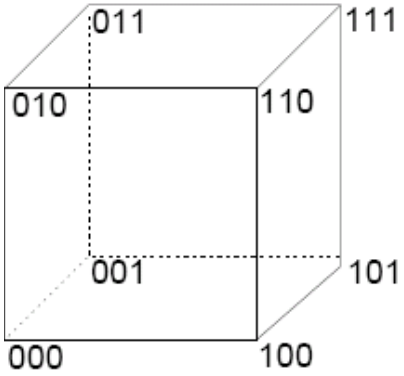


Example: *E. coli* fitness landscape

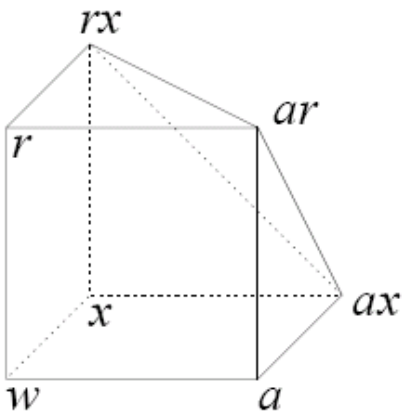
- 9 mutations in 3 groups: $(a, b, c), (r, s, t), (x, y, z)$, 37 genotypes

w	a	b	c	r	s	t	x	y	z
a				ar	as	at	ax	ay	az
b				br	bs	bt	bx	by	bz
c				cr	cs	ct	cx	cy	cz
r	ar	br	cr				rx	ry	rz
s	as	bs	cs				sx	sy	sz
t	at	bt	ct				tx	ty	tz
x	ax	bx	cx	rx	sx	tx			
y	ay	by	cy	ry	sy	ty			
z	az	bz	cz	rz	sz	tz			

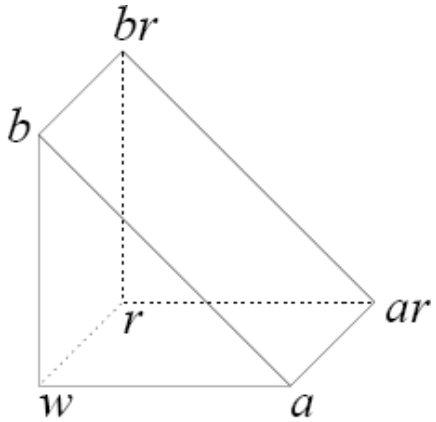
Three locus subsystems



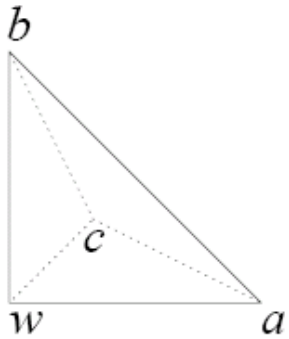
74



16

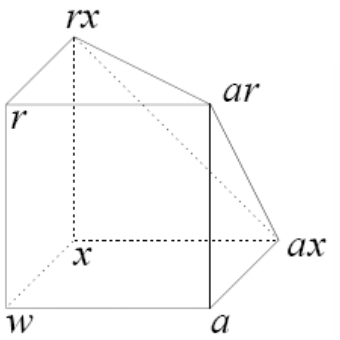


6



1

Secondary polytope



- Connected shapes differ by the sign of exactly one gene interaction, e.g., 36 and 37 by $r \cdot ax - x \cdot ar$.

Markov basis

- The experimental design suggests considering the minimal Markov basis of the interaction space. It contains 27 “standard” interactions, plus 216 additional “non-standard” interactions:

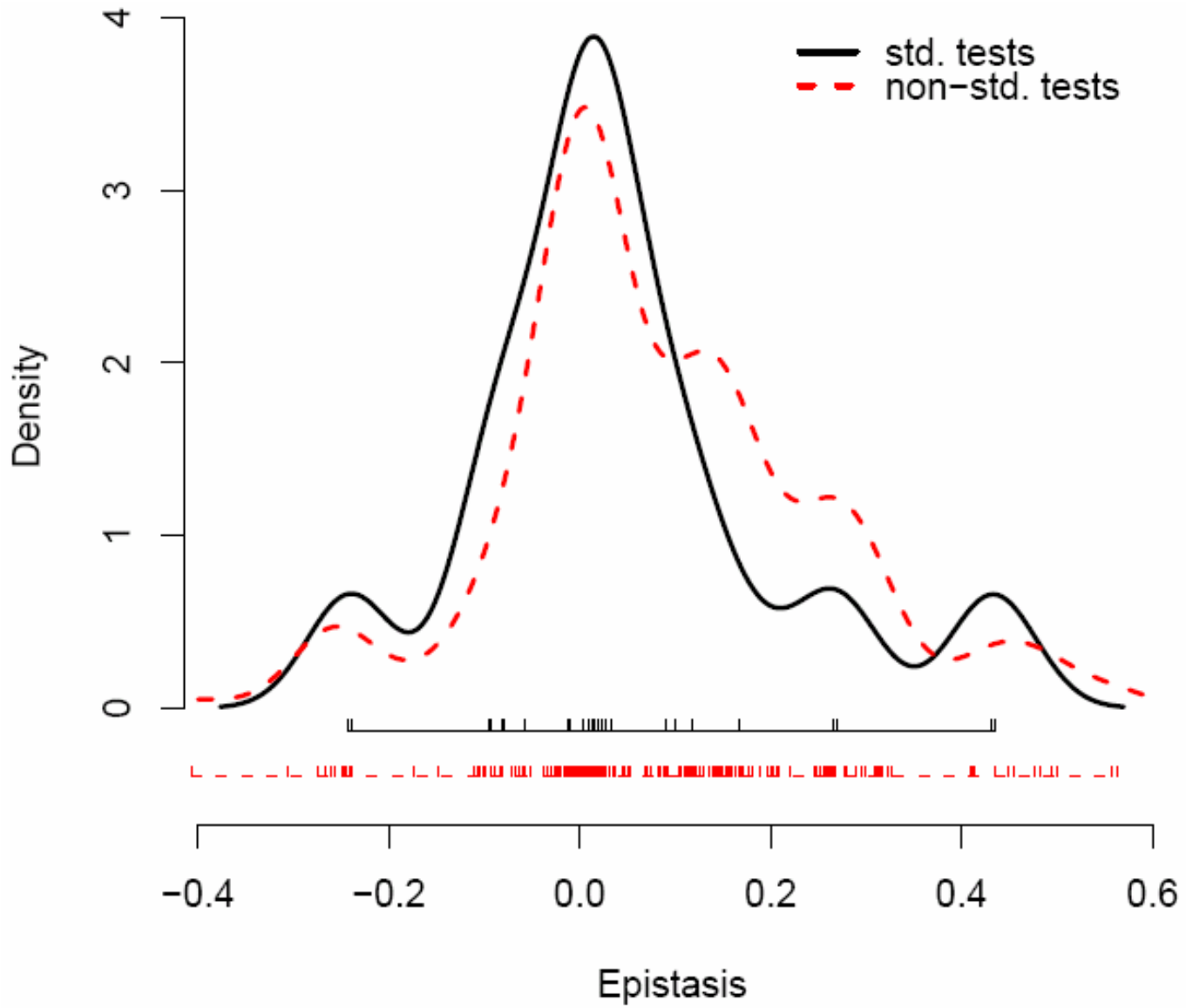
$$w \cdot ar - a \cdot r \qquad (27)$$
$$ar \cdot bs - as \cdot br \qquad (108)$$
$$a \cdot br - b \cdot ar \qquad (108)$$

w	a	b	c	r	s	t	x	y	z
a				ar	as	at	ax	ay	az
b				br	bs	bt	bx	by	bz
c				cr	cs	ct	cx	cy	cz
r	ar	br	cr				rx	ry	rz
s	as	bs	cs				sx	sy	sz
t	at	bt	ct				tx	ty	tz
x	ax	bx	cx	rx	sx	tx			
y	ay	by	cy	ry	sy	ty			
z	az	bz	cz	rz	sz	tz			

$$\rho : \Delta_G \rightarrow \Pi_G$$

- In all tests, allele frequencies are fixed.

Standard and non-standard tests



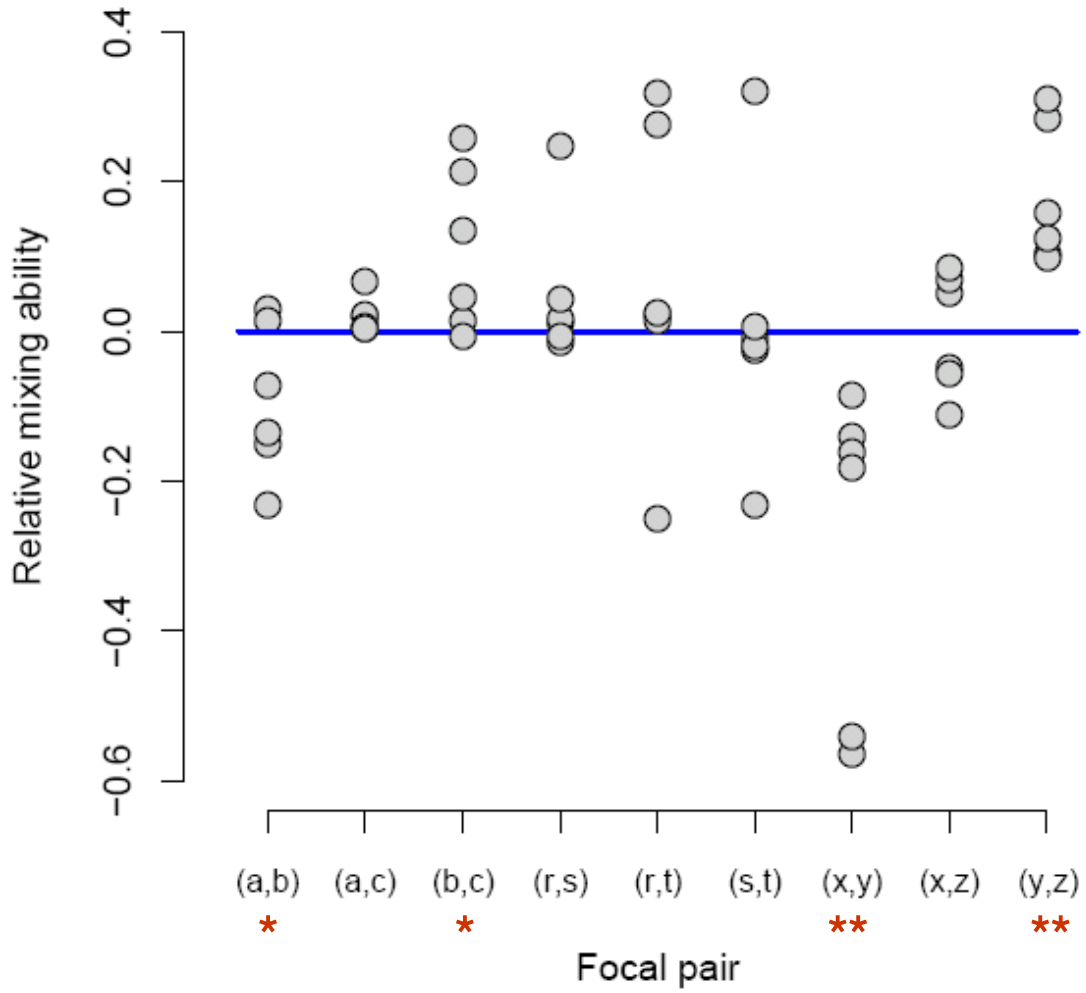
Some non-standard tests reveal mixing ability

$a \cdot br - b \cdot ar$

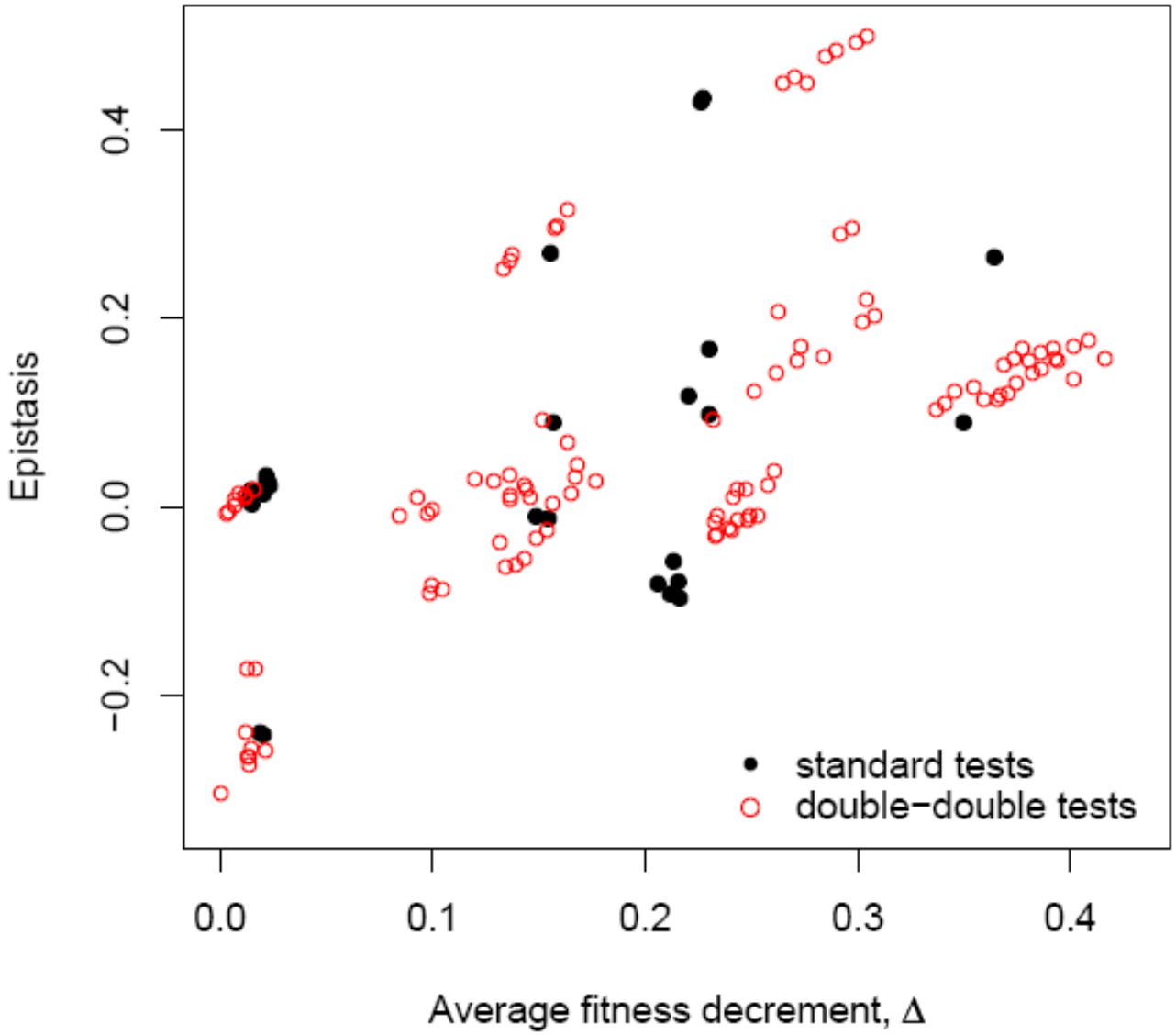
$a \cdot bs - b \cdot as$

…

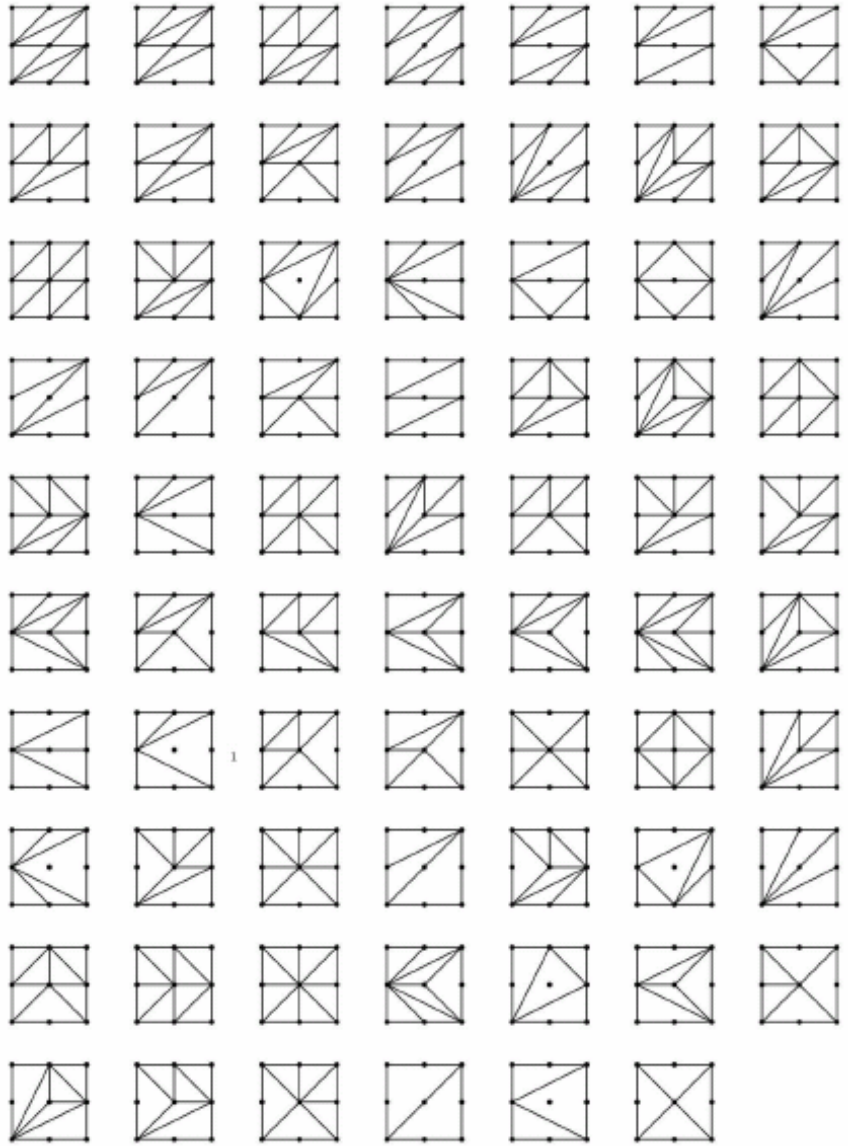
test the relative mixing ability of mutation a versus b



Compensatory mutations

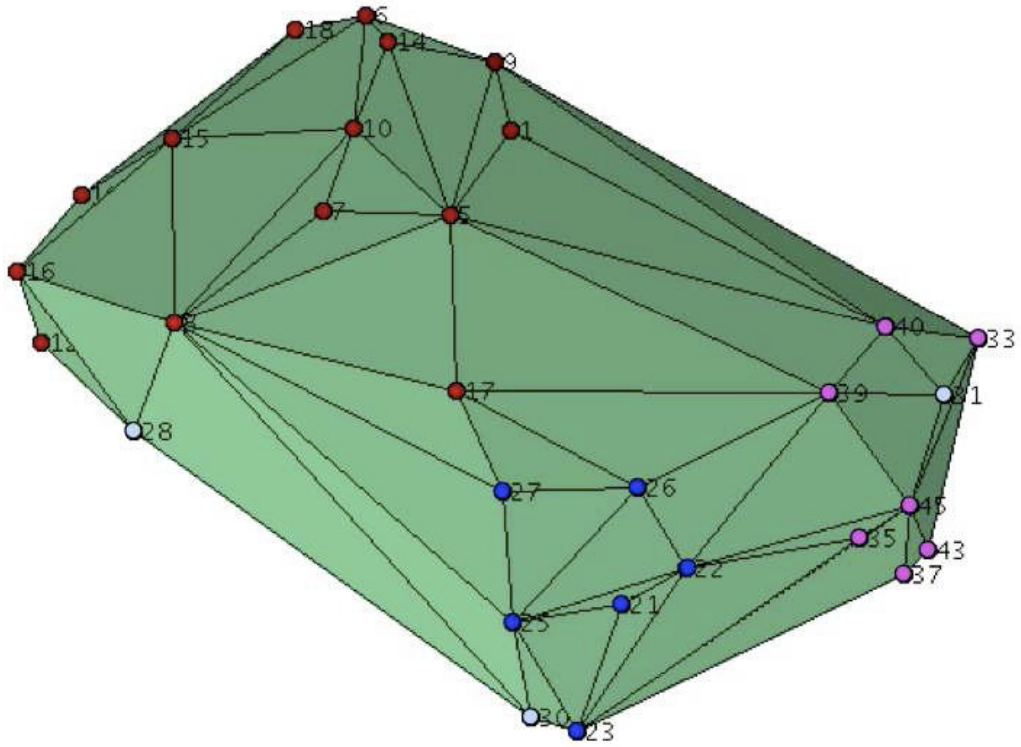


More shapes: diploids



I. Hallgrimsdottir, Debbie Yuster:
A complete classification of epistatic two-locus models

More shapes: human SNPs



Peter Huggins, L. Pachter, B. Sturmfels:
Towards the human genotope

References

- Beerenwinkel N, Pachter L, Sturmfels B
  Epistasis and shapes of fitness landscapes
  Statistica Sinica
- Beerenwinkel N, Pachter L, Sturmfels B, Elena SF, Lenski RE
  Analysis of epistatic interactions and fitness landscapes using a new geometric approach
  BMC Evolutionary Biology