



University of Zurich UZH



**Wagner Lab**

Dr. Tess Brewer  
 Dr. Shradha Karve  
 Dr. Pouriya Dasmeh  
 Camille Jourdan  
 Dr. Andrei Papkou  
 Dr. Diego Pesce  
 Michael Rechsteiner  
 Aneette Schmid  
 Dr. Bharat Ravi Iyengar  
 Michael Schmutzner  
 Dr. Jordi van Gestel  
 Dr. Teresa Vaello  
 Caua Westmann  
 Dr. Jia Zheng

**Recent collaborators**

Martin Ackermann (ETHZ)  
 Andrew Barbour (UZH)  
 Florian Altermatt (UZH)  
 Brett Calcott (Univ. of Sydney, Australia)  
 Mario Fares (Trinity College, Ireland)  
 Miguel Fortuna/Jordi Bascompte (UZH)  
 Laurent Hébert-Dufresne (University of Vermont)  
 Mirna Galeic (SFI)  
 Michael Hochberg (SFI, Montpellier)  
 Rolf Kummerli (UZH)  
 Eric Libby (SFI, Umea University)  
 Robert Maxfield (The Santa Fe Institute, USA)  
 Olivier Martin (Orsay, France)  
 Scott Ortman (Univ. of Colorado, Boulder)  
 Paul Sniegowski (Univ. Pennsylvania)



European Research Council



FONDS NATIONAL SUISSE  
 SCHWEIZERISCHER NATIONALFONDS  
 FONDO NAZIONALE SVIZZERO  
 SWISS NATIONAL SCIENCE FOUNDATION

1



2



3

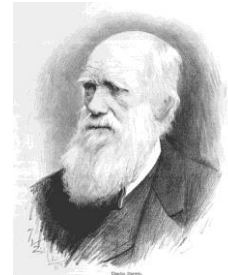
1859: Charles Darwin and «The origin of species»

Natural selection as a central cause of life's diversity

All life originated from a common ancestor

One main area of ignorance:

*"The laws governing inheritance are for the most part unknown."*

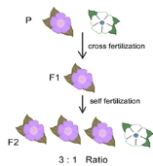


4

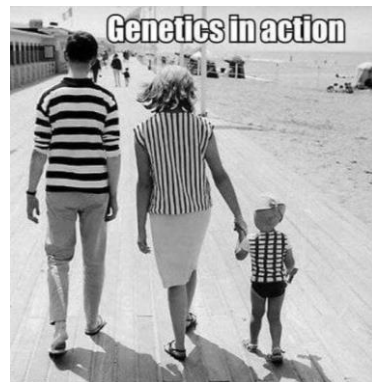
1866: Gregor Mendel

Solved part of the inheritance problem

Showed that traits can behave like indivisible particles when inherited



5

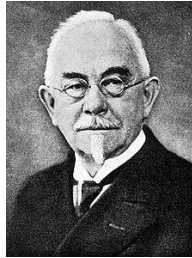


6

1909: Wilhelm Johannsen

Danish botanist who coined three key concepts

- Gene**  
Mendel's particulate units of inheritance  
(still completely unclear what genes were)
- Genotype**  
in modern language: all of an organism's DNA
- Phenotype**  
All other observable traits, including fitness



7

1930-1940s: The modern synthesis

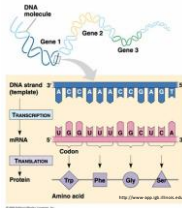
- Named after a 1942 book by Julian Huxley
- Main characters: R.A. Fisher, Sewall Wright, J.B.S. Haldane
- Joined three elements
  - The importance of natural selection
  - Mendel's laws of particulate inheritance
  - Population thinking
- Created population genetics
- Quantitative and mathematical** theory that predicts how selection changes a population
- Challenging for many biologists



8

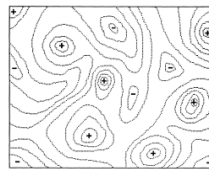
1944-: The molecular era

- 1944: Oswald Avery discovers that DNA is the material basis of inheritance
- 1953: James Watson and Francis Crick find the chemical structure of DNA
- 1950s-1960s: transcription, translation, and the genetic code



9

Wright developed the adaptive landscape to explain essential concepts of population genetics in non-mathematical language

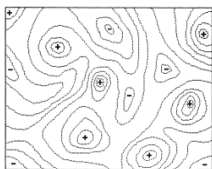


Location: genotype  
Elevation: fitness

Wright, Proc. 6th Int. Congr. Genet., 1932

10

Wright developed the adaptive landscape to explain essential concepts of population genetics in non-mathematical language



**Evolving population:** multiple organisms that explore an adaptive landscape through mechanisms that

- 1) create variation  
mutation  
recombination
- 2) cause differential survival/reproduction  
natural selection  
genetic drift

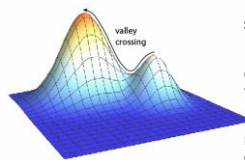
**Selection drives evolving populations uphill**

Wright, Proc. 6th Int. Congr. Genet., 1932

11

11

Wright developed the adaptive landscape to explain essential concepts of population genetics in non-mathematical language



**Selection drives evolving populations uphill**

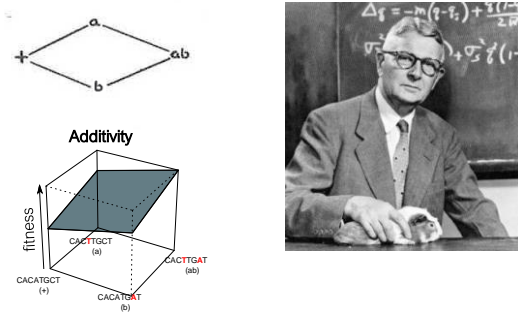
On a multi-peaked landscape «things may need to get worse before they can become better»

Finding the highest peaks (best-adapted organisms) on a multi-peaked landscape cannot be accomplished by selection alone

12

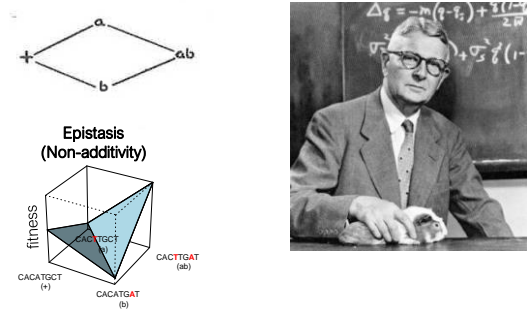
12

Wright realized that mutations can interact in complicated ways



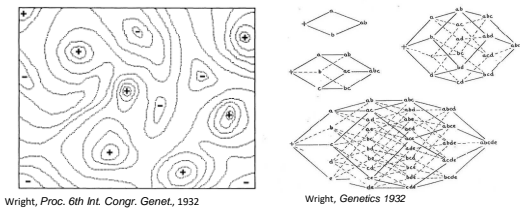
13

Wright realized that mutations can interact in complicated ways



14

Epistasis can help create multi-peaked landscapes



15

Adaptive landscapes are vast and multi-dimensional

A DNA sequence (gene) with 150 base pairs:  
 $4^{150} \approx 10^{90}$  possible genotypes

A protein sequence with 150 amino acids:  
 $20^{150} \approx 10^{195}$  possible genotypes (amino acid sequences)

Compare:  $\approx 10^{80}$  hydrogen atoms in the universe

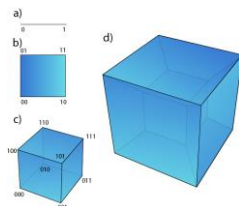
16

Adaptive landscapes are vast and multi-dimensional

Simplified model of genotype: binary string  
 (...0101110101011...)

The 'space' of all possible genotypes becomes a hypercube

Each axis corresponds to one landscape dimension



17

The landscape metaphor can be interpreted in multiple ways

Location	Elevation
genotype	fitness
genotype	any (scalar) phenotype
phenotypic traits (one per dimension)	fitness

18

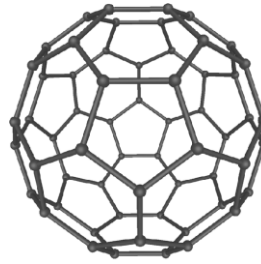
Landscape thinking influences many areas of science

Examples:

- chemistry
- operations research
- psychology

19

Buckminsterfullerenes are complex  $C_{60}$  carbon-cages that form in the interstellar medium

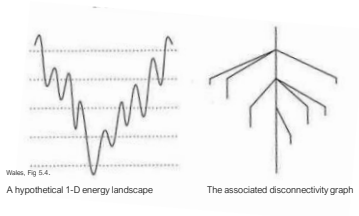


truncated icosahedron  
20 hexagons  
12 pentagons  
No pentagons are adjacent



20

Positions of atoms in a molecule can be described by a potential energy landscape



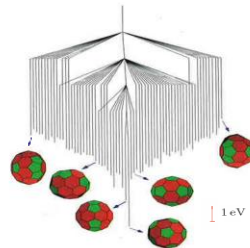
Wales, Fig 5.4.  
A hypothetical 1-D energy landscape

The associated disconnectivity graph

**Axes:** lengths of chemical bonds  
**Elevation:** potential energy (low energy – high stability)  
**Variation/generation:** thermal energy  
**Minima:** stable configurations

21

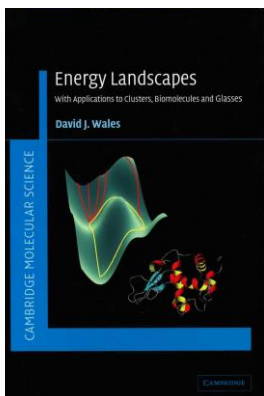
Buckminsterfullerenes have a complex energy landscape



The disconnectivity graph shows multiple minima near the global minimum

Wales, Fig 8.35.

22



23

The traveling salesman problem (TSP) is of immense practical and theoretical importance

Task: identify the shortest journey ("tour") that visits each of  $n$  locations exactly once

Important in operations research  
energy and time-efficient travel  
efficient chip design  
minimizing time to move a telescope through various positions  
...

benchmark problem for combinatorial optimization



A secretary planning a salesman's route with the 'pin and string' method in the 1920s

Cook, Fig. 2.3.

24

The traveling salesman problem (TSP) is of immense practical and theoretical importance

A 13,509 US cities tour first solved in 1996



Cook, Fig. 8.8.

25

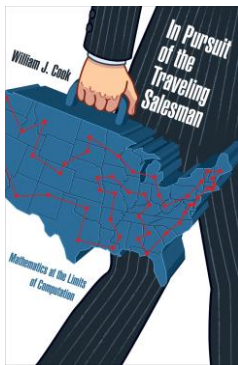
The traveling salesman problem (TSP) is of immense practical and theoretical importance

Location: a journey

Elevation: length of the tour



26



27

Many researchers consider human creativity a Darwinian process

"a seething caldron of ideas, where ... treadmill routine is unknown, and the unexpected seems the only law...the genius of discovery depends altogether on the number of these **random notions and guesses** which visit the investigator's mind."

William James, 1890  
American philosopher and psychologist

"I have to compare myself with an Alpine climber, who, not knowing the way, ascends stowly and with toil, and is often compelled to retrace his steps because his progress stopped; sometimes by reasoning, and sometimes by accident, he hits upon traces of a fresh path, which again leads him a little further..."

Hermann von Helmholtz, 1891  
physician and physicist

28

Many researchers consider human creativity a Darwinian process

*Psychological Review*  
1960, Vol. 67, No. 6, 380-400

**BLIND VARIATION AND SELECTIVE RETENTION  
IN CREATIVE THOUGHT AS IN OTHER  
KNOWLEDGE PROCESSES'**

DONALD T. CAMPBELL

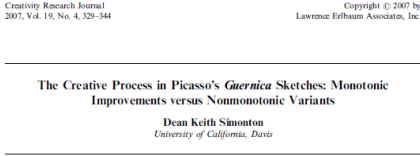
29

Pablo Picasso's Guernica



30

A study on Picasso's process in creating Guernica

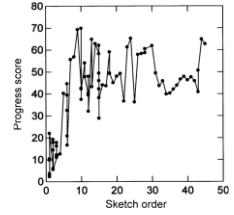


31

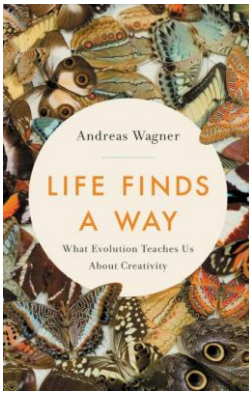
A study on Picasso's process in creating Guernica

Picasso drew 45 sketches of Guernica  
 Sketches are temporally ordered from early to late  
 Simonton shuffled the sketches randomly  
 Asked five judges to evaluate their similarity to Guernica  
 Plotted similarity against temporal order

**Picasso's process does not involve simple hill climbing**



32



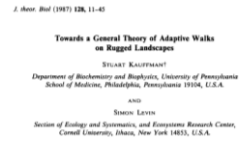
33

A worst-case scenario: Adaptation on a maximally rugged, "uncorrelated" landscape

Biopolymers (DNA, protein) of length N  
 B possible monomers  
 B=4 for DNA  
 B=20 for proteins  
 B possible genetic variants ("alleles") at each position ("locus") in the string

Example:  
 TAGCTCT  
 TAGAACT

Two alleles at locus 5

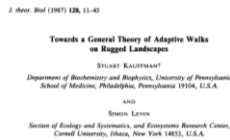


$B^N$  possible genotypes  
 Each genotype has  $N(B-1)$  1-mutant neighbors

34

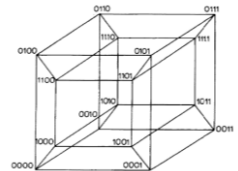
A worst-case scenario: Adaptation on a maximally rugged, "uncorrelated" landscape

Assign fitness values randomly to each of the  $B^N$  possible strings from a some distribution on a real interval (x,y)  
 Rank all fitness values  
 Replace fitness values with their rank



A worst-case scenario: Adaptation on a maximally rugged, "uncorrelated" landscape

Simple special case: B=2  
 $2^N$  possible genotypes  
 Assign integers in  $(1, 2^N)$  randomly with uniform distribution to vertices of the hypercube



35

36

A worst-case scenario: Adaptation on a maximally rugged, "uncorrelated" landscape

Neighboring genotypes have uncorrelated fitness values

Number of local peaks below the single global peak

$$\frac{B^N}{N(B-1) + 1}$$

Scales exponentially in  $N$

37

A worst-case scenario: Adaptation on a maximally rugged, "uncorrelated" landscape

Adaptive random walks step from a genotype to a randomly chosen fitter neighbors with equal probability

Number of steps before such a walk can go no further (has reached a local peak)?

$$\log_2 D = \log_2(N(B-1)) \quad (\text{upper bound})$$

Increases much more slowly than  $N$  or  $B$

Reason: number of peaks increases exponentially

Example:  
 $N=10, B=20: \log_2 D = 7.6$   
 $N=100, B=20: \log_2 D = 10.9$

38

A worst-case scenario: Adaptation on a maximally rugged, "uncorrelated" landscape

Number of peaks that can be reached from the genotype with lowest fitness

$$D^{(\log_2 D - 1)/2} = (N(B-1))^{(\log_2 N(B-1) - 1)/2} \quad (\text{upper bound})$$

Example:  $N=100, B=20$

$1.6 \times 10^{16}$  accessible peaks

Total number of peaks  $6.7 \times 10^{126}$

Fraction accessible:  $2.4 \times 10^{-111}$

39

Popular theoretical models of adaptive landscapes

1. Uncorrelated model

a.k.a the "house of cards" model

every mutation can have large effects and thus destroy the "house of cards" of a complex biological system



40

Popular theoretical models of adaptive landscapes

2. Additive model

- $g$  ... genotype
- $g_i$  ... genotype at position (locus)  $i$
- $a_i(g)$  ... real-valued function of  $g_i$
- $w(g)$  ... fitness of genotype

$$w(g) = \sum_{i=1}^N a_i(g_i)$$

fitness effect of any one mutation is independent of genotype at other loci  
 single-peaked landscape



41

Popular theoretical models of adaptive landscapes

3. NK-Model

- $N$  loci
- Fitness contribution of each locus depends on allele at the locus itself
- alleles at  $K$  other loci

Two key parameters:  $N$  and  $K$

Other parameters:  
 Distribution of the  $K$  loci that influence locus  $i$  (e.g., random and uniform, adjacent, etc.)  
 Number of monomers  $B$   
 Distribution of fitness values to assign to each locus and allele

42

Popular theoretical models of adaptive landscapes

3. NK-Model

Fitness contribution of locus  $i$  to genotype  $g$

$$w_i(g) = w_i(g_i, g_{j_1}, \dots, g_{j_k})$$

$B^{k+1}$  possible values of this argument  
choose as the function values  $B^{k+1}$  random numbers on the interval (0,1)

Fitness of genotype  $g$

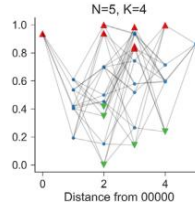
$$w(g) = 1/N \sum_{i=1}^N w_i(g)$$

43

Popular theoretical models of adaptive landscapes

3. NK-Model

special case  $K=N-1$ : uncorrelated model



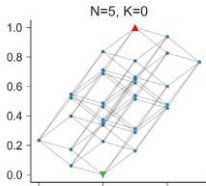
Obolski et al 2018 Rep. Prog. Phys. 81

44

Popular theoretical models of adaptive landscapes

3. NK-Model

special case  $K=0$ : additive model



Obolski et al 2018 Rep. Prog. Phys. 81

45

Popular theoretical models of adaptive landscapes

3. NK-Model

$0 < K < N-1$ : very complicated landscape structure

few analytical results

"tunable ruggedness"

46

Popular theoretical models of adaptive landscapes

4. Rough Mount Fuji model

$$w(g) = c\eta(g) - d(g, g_p)$$

$g_p$ ...reference (peak) genotype  
 $d(g, g_p)$ ... (Hamming) distance to reference genotype  
 $c$ ...constant weight factor  
 $\eta(g)$ ...random variable, uncorrelated among genotypes

$c$  large: uncorrelated  
 $c$  small: single-peaked

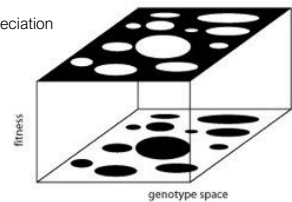
47

Popular theoretical models of adaptive landscapes

5. Holey adaptive landscapes

Only fitness one or zero

Developed to study speciation



Gavrilets 1997 TREE 12: 307

48



Popular theoretical models of adaptive landscapes

6. Models with biophysically/biochemically realistic genotype-phenotype relationships



Modelling neutral and selective evolution of protein folding

DAVID J. LIPMAN AND W. JOHN WILBUR  
Proc. R. Soc. Lond. B (1991) 245, 7-11



From sequences to shapes and back: a case study in RNA secondary structures

PETER SCHUSTER<sup>1,2</sup>, WALTER FONTANA<sup>1</sup>, PETER F. STADLER<sup>1,3</sup> AND IVO L. HOFACKER<sup>2</sup>  
Proc. R. Soc. Lond. B (1994) 255, 279-284



Evolutionary Plasticity and Innovations in Complex Metabolic Reaction Networks

João F. Matias Rodrigues<sup>1,2</sup>, Andreas Wagner<sup>1,2,3</sup> PLoS Computational Biology  
December 2009 | Volume 5 | Issue 12 | e1000613

49

Quantifiers of landscape ruggedness

2. (Semi)local quantifiers

Fitness autocorrelation

Along an ensemble of random walks  $\{g_t\}$  on the landscape, compute

$$R(t, s) = \frac{M(w(g_t)w(g_{t+s})) - M(w(g_t))M(w(g_{t+s}))}{V(w(g_t))}$$

(M, V: mean, variance)

51

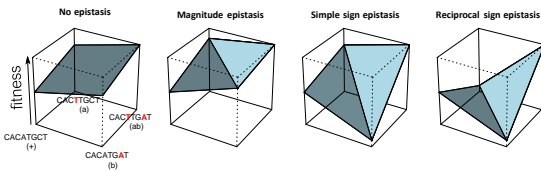
Quantifiers of landscape ruggedness

2. (Semi)local quantifiers

Incidence of pairwise epistasis

Reciprocal sign epistasis is necessary for the existence of multiple peaks

Poelwijk et al., JTB 2010



53

Quantifiers of landscape ruggedness

1. Global quantifiers

Number of peaks

Peak accessibility

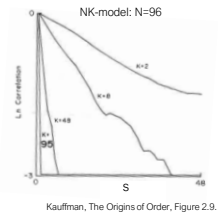
starting from a given genotype (randomly chosen, low fitness etc.) determine the fraction of all (shortest) paths to a given peak (global, high-fitness local etc.) that are monotonically fitness-increasing

50

Quantifiers of landscape ruggedness

2. (Semi)local quantifiers

Fitness autocorrelation  
More rapid decay in more rugged landscapes



Kauffman, The Origins of Order, Figure 2.9.

52

Quantifiers of landscape ruggedness

2. (Semi)local quantifiers

epistasis of 3<sup>rd</sup>, 4<sup>th</sup> etc. order  
various approaches

Fourier analysis (Stadler, J. Math. Chem 1996)  
Walsh transform (Weinreich et al., Curr. Opin. Genet. Dev 2013)

higher order epistasis may be sparse

Affecting few loci/alleles (Poelwijk et al., Nat. Comm. 2019)

54

# Empirical landscapes

## Ammonoid swimming efficiency

Antibiotic resistance evolution  
 TEM-1 beta-lactamase  
 Dihydrofolate reductase

Ribozymes

Gene regulation

55

## An adaptive landscape of Ammonoid shell shape

Ammonoids (aka Ammonites)  
 are extinct molluscs

Closely related to today's Nautilus

Multichambered spiral shell

Rich fossil record



56

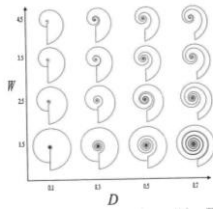
## An adaptive landscape of Ammonoid shell shape

Ammonoids had diverse shell shapes

Their *morphospace* can be described  
 by two parameters

W...whorl expansion rate  
 diameter of shell opening  
 relative to distance of coiling  
 axis to inner edge of  
 shell opening

D...distance from coiling axis to  
 inner edge of the shell aperture



McGhee 2007, The Geometry of Evolution, Figure 4.5.

57

## An adaptive landscape of Ammonoid shell shape

Chamberlain built plexiglass models of  
 Ammonoids with various shell shapes

Dragged them through a water tank

Measured drag coefficient  
 a measure of energy expenditure  
 when swimming



John Chamberlain, Paleontology 1976.

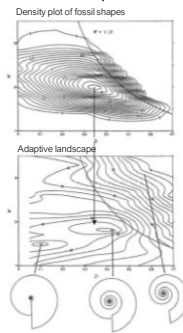
58

## An adaptive landscape of Ammonoid shell shape

The adaptive landscape has two peaks

Fossils cluster preferentially around  
 one peak

Both peaks are occupied  
 discovered later through new fossil finds:  
 Saunders et al., Paleobiology 2004



McGhee 2007, The Geometry of Evolution, Figure 5.1.

59

# Empirical landscapes

Ammonoid swimming efficiency

Antibiotic resistance evolution  
**TEM-1 beta-lactamase**  
 Dihydrofolate reductase

Ribozymes

Gene regulation

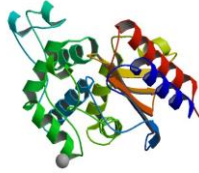
60

TEM-1 beta-lactamase is a model protein for the evolution of antibiotic resistance

Causes resistance against the antibiotic ampicillin

Five DNA mutations suffice to create strong resistance against cefotaxime

Cefotaxime is a WHO «essential medicine»

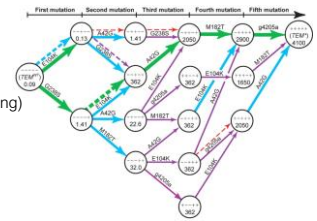


TEM-1 beta-lactamase is a model protein for the evolution of antibiotic resistance

Weinreich et al., created all 32 combinations of the five point mutations (not all shown here)

5! = 120 Shortest paths to the cefotaxime «peak»

Fewer than 20% accessible (cefotaxime resistance increasing)

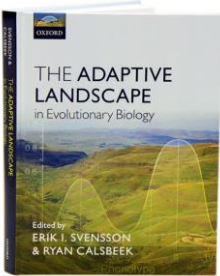


Weinreich et al., Science 2006

61

62

A renaissance of work on landscapes



**Empirical fitness landscapes and the predictability of evolution**  
J. Arjan C.M. de Visser and Joachim Krup *NATURE REVIEWS | GENETICS*

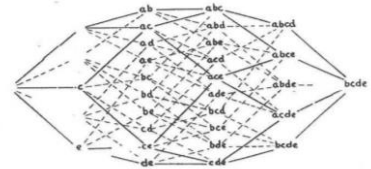
**Local fitness landscape of the green fluorescent protein** *NATURE*  
Dimitris A. Katsouris<sup>1,2,3,4</sup>, George V. Katsouris<sup>1,2,3,4</sup>, George V. Katsouris<sup>1,2,3,4</sup>, George V. Katsouris<sup>1,2,3,4</sup>, George V. Katsouris<sup>1,2,3,4</sup>, George V. Katsouris<sup>1,2,3,4</sup>, George V. Katsouris<sup>1,2,3,4</sup>, George V. Katsouris<sup>1,2,3,4</sup>, George V. Katsouris<sup>1,2,3,4</sup>, George V. Katsouris<sup>1,2,3,4</sup>

**The fitness landscape of a tRNA gene** *SCIENCE*  
Chuan Li, Weidong Qian, Caimen J. Markham, Haohai Zhang

Combinatorially complete landscape data is important for landscape analysis

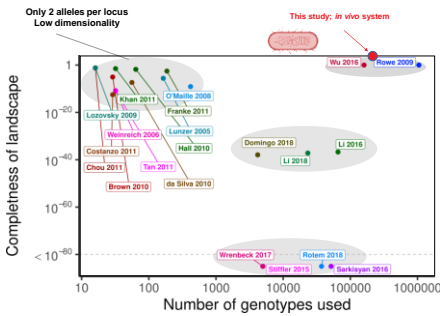
Landscape data is combinatorially complete if the following condition holds for any two genotypes whose phenotype (fitness) has been quantified:

The phenotype (fitness) has also been quantified for all genotypes that lie on each shortest path between the two genotypes



63

64



Andrei Pakkou

## Empirical landscapes

Ammonoid swimming efficiency

Antibiotic resistance evolution  
TEM-1 beta-lactamase  
**Dihydrofolate reductase**

Ribozymes

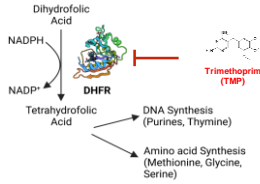
Gene regulation

65

66

Dihydrofolate reductase (DHFR) is a key enzyme in 1-carbon metabolism

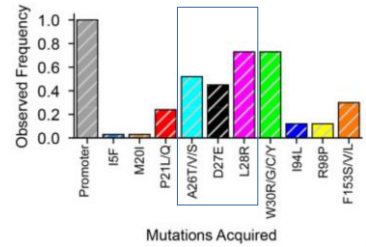
Produces essential building blocks for DNA and RNA  
 Is inhibited by antibiotics (trimethoprim) and antifolates (anti-cancer drugs, e.g., methotrexate)  
 Cells evolve resistance against trimethoprim through mutations in DHFR



Andrei Papkou

67

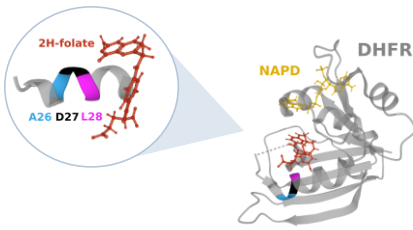
Mutations at several consecutive amino acids are frequently involved in the evolution of TMP resistance



Tamer, MBE 2019

68

Several of these mutations occur at the dihydrofolate binding site



Andrei Papkou

69

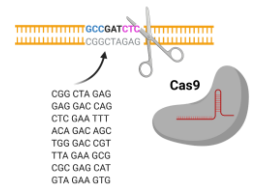
CRISPR-Cas genome editing can help to create large libraries of DHFR variants

Produce all 4<sup>9</sup>= 262,144 possible variants at 9 consecutive nucleotides encoding A26, D27, L28

Combinatorially complete library

On the *E.coli* chromosome *in vivo*

Create a large population of *E.coli* cells in which different cells harbor different variants

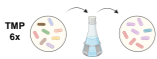


Andrei Papkou

70

Mass-selection and deep sequencing can help measure the fitness of all variants in one experiment

1. Mass library selection



2. Deep sequencing before and after selection



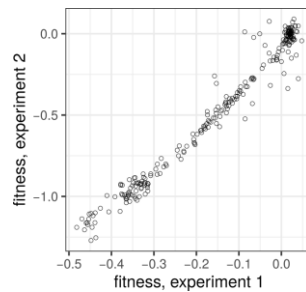
3. Genotype counting

genotype	$n_{x,pre}$	$n_{x,post}$
TCGTTGCTT	235	10
TGTGTAAA	120	35
CGTATGCA	587	987
...		
GGTGGTGC	224	39

$n_{x,pre}$  ... sequence reads with genotype x before selection  
 $n_{x,post}$  ... sequence reads genotype x after selection

fitness of genotype x:  $w_x \sim n_{x,post} / n_{x,pre}$

Fitness measurements are highly reproducible between experiments

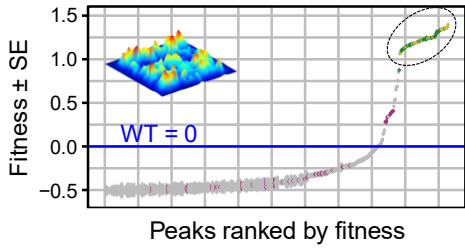


Andrei Papkou

71

72

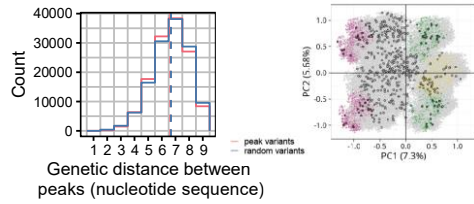
The DHFR adaptive landscape harbors 73 fitness peaks far above the wild-type for antibiotic resistance



Andrei Papkou

73

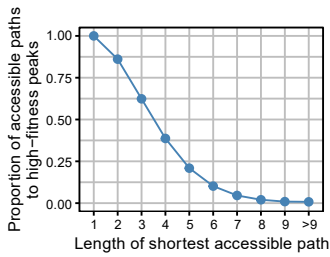
The landscape's peaks are scattered through genotype space



74

Andrei Papkou

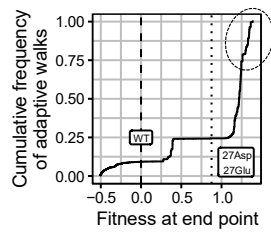
High fitness peaks are accessible even from very distant genotypes



Andrei Papkou

75

Most evolving populations reach a high fitness peak

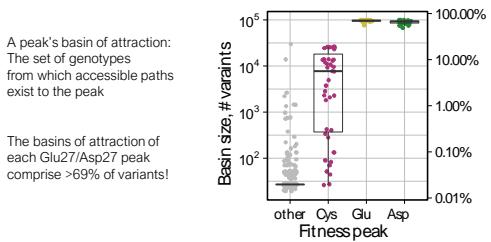


76% of evolving populations reach a high fitness peak  
Based on 10<sup>6</sup> adaptive walks starting from random genotypes

Andrei Papkou

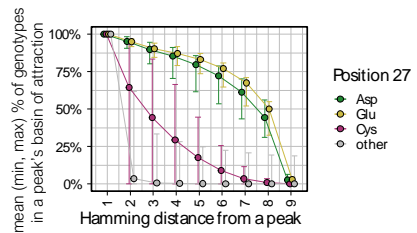
76

High fitness peaks have very large basins of attraction



77

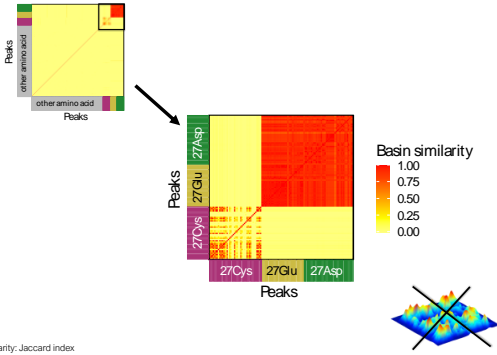
Many genotypes in a peak's basin of attraction are far away from the peak



The majority of variants in a basin also have low fitness (not shown)

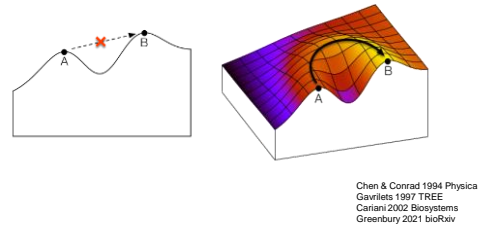
78

The basins of attraction of high peaks overlap greatly



79

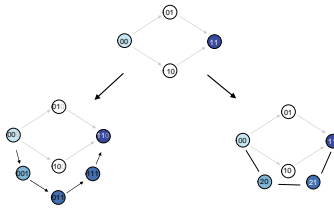
Extradimensional by-passes can help overcome fitness valleys



Andrei Papkou

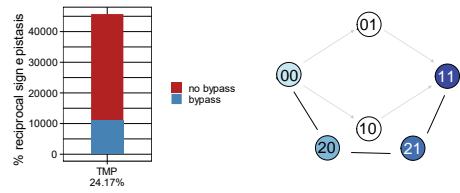
80

Extradimensional by-passes can take various forms



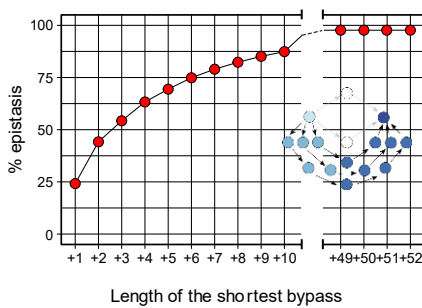
81

25% of fitness dips created by reciprocal sign epistasis can be by-passed with only one extra step.



82

Longer by-passes can help overcome most fitness dips.



83

Summary

The DHFR antibiotic resistance landscape has many peaks

Despite this ruggedness, high fitness peaks are reached by most evolving populations

Bendtsen et al. PLoS Biology (2019)

84

# Empirical landscapes

Ammonoid swimming efficiency

Antibiotic resistance evolution  
 TEM-1 beta-lactamase  
 Dihydrofolate reductase

## Ribozymes

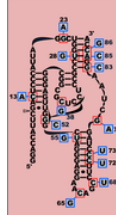
Gene regulation

85

## Adaptive landscapes of two ribozymes (RNA enzymes)

### Hepatitis D virus (HDV) ribozyme

natural self-cleaving



### Class II ligase

synthetic ligates an RNA molecule to itself



Bendtsen et al. PLoS Biology (2019)

86

## Adaptive landscapes of two RNA enzymes

Start from one reference HDV and one reference ligase ribozyme they differ at 14 sites

Create all  $2^{14}=16384$  possible variants

Measure "fitness" of each variant *in vitro*  
 ability to self-cleave relative to a reference  
 ability to ligate relative to a reference

Study landscape with respect to both phenotypes

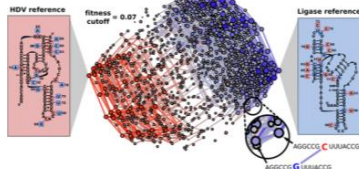


Figure shows only genotypes with fitness >0.07 x reference  
 Node color: dominant activity, node size: activity relative to reference

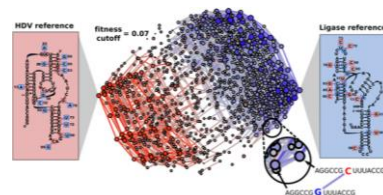
Bendtsen et al. PLoS Biology (2019)

87

## Adaptive landscapes of two RNA enzymes

The HDV landscape is much more rugged than the ligase landscape  
 982 peaks vs. 68 peaks

More than 50% of genotypes (9032) can catalyze both reactions although at very different rates



Bendtsen et al. PLoS Biology (2019)

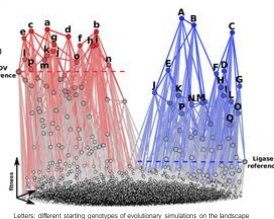
88

## Adaptive landscapes of two RNA enzymes

Simulate evolution from a HDV sequence towards a ligase sequence

Assumptions:

- It would be beneficial for an organism to be capable of catalyzing both the cleavage and the ligase reaction
- A gene duplication has created two copies of the gene encoding the self-cleaving HDV enzyme  
 Allows the preservation of self-cleavage ability while 'searching' for genotypes with ligase activity



Letters: different starting genotypes of evolutionary simulations on the landscape

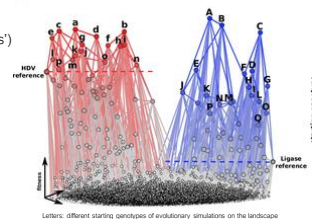
Bendtsen et al. PLoS Biology (2019)

89

## Adaptive landscapes of two RNA enzymes

Survival probability of a sequence proportional to ligase activity ('fitness')

Various population sizes, mutation rates, starting genotypes



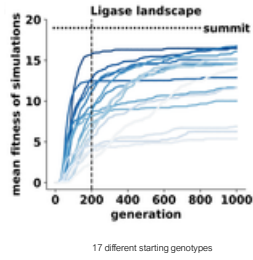
Letters: different starting genotypes of evolutionary simulations on the landscape

Bendtsen et al. PLoS Biology (2019)

90

Adaptive landscapes of two RNA enzymes

Success of adaptive evolution depends on the initial genotype  
Some populations arrive close to the 'summit' (highest fitness genotype)



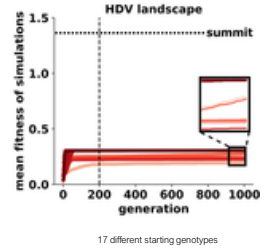
Bendixen et al. PLoS Biology (2019)

91

Adaptive landscapes of two RNA enzymes

Simulate evolution from a ligase sequence towards a HDV sequence

All populations get stuck far below the summit



Bendixen et al. PLoS Biology (2019)

92

Summary

- Landscape ruggedness varies among RNA molecules
- Landscape ruggedness matters for adaptive evolution
- Ribozyme promiscuity facilitates adaptive evolution

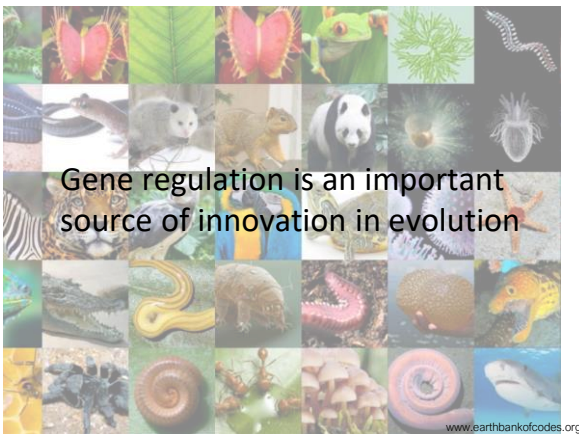
Bendixen et al. PLoS Biology (2019)

93

Empirical landscapes

- Ammonoid swimming efficiency
- Antibiotic resistance evolution
  - TEM-1 beta-lactamase
  - Dihydrofolate reductase
- Ribozymes
- Gene regulation**

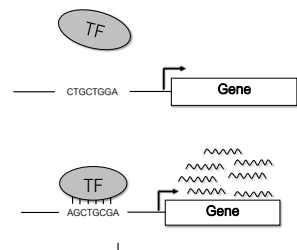
94



95

Gene regulation involves proteins called transcription factors (TFs)

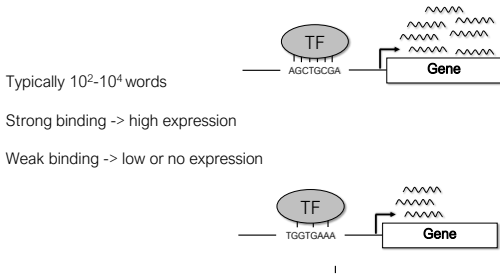
They bind short DNA words near a gene  
Binding enables transcription (gene expression)



96



Any one TF can bind multiple DNA words



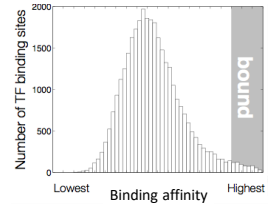
97

Large scale measurements allow the analysis of affinity landscapes

Protein binding microarrays measure the binding strength (affinity) for any one TF to all  $4^8=65536$  DNA words of length 8.

Only a small subset is bound with high affinity

Construct affinity landscape from these sequences



Berger et al., Nat. Biotechnol., 2006  
Berger et al., Nat. Protoc., 2009

Joshua L. Payne

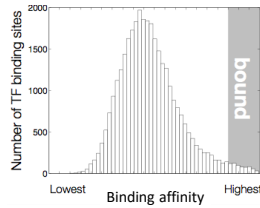
98

98

Large scale binding affinity measurements allow the analysis of affinity landscapes

Data available for >1000 transcription factors (>100 species)

>1000 landscapes



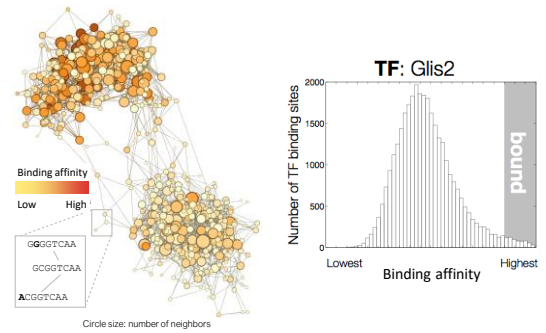
Berger et al., Nat. Biotechnol., 2006  
Berger et al., Nat. Protoc., 2009

Joshua L. Payne

99

99

The affinity landscape of the mouse TF Glis2

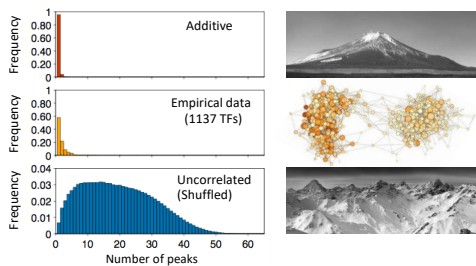


Joshua L. Payne

Payne & Wagner, Science 2014

100

Landscape of binding affinity have one to few peaks.



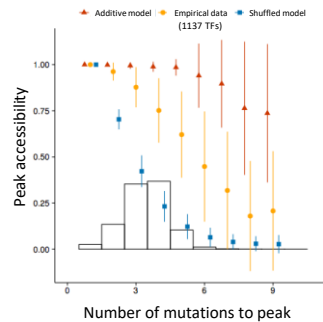
Joshua L. Payne

101

Aguilar, Payne, Wagner, Nat Ecol. Evol., 2017

101

Peaks are highly accessible via affinity-increasing paths.

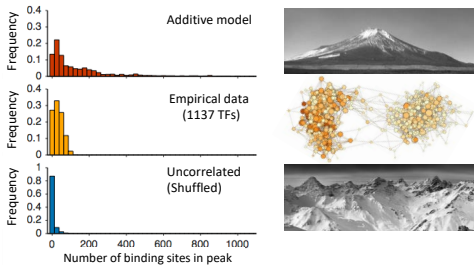


102

Aguilar, Payne, Wagner, Nat Ecol. Evol., 2017

102

Landscape peaks often comprise many binding sites.



Joshua L. Payne

103

Aguilar, Payne, Wagner, *Nat Ecol. Evol.*, 2017

103

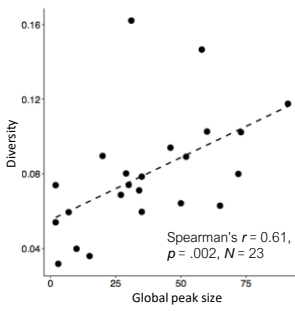
Binding sites from broad peaks should exhibit more genetic diversity than binding sites from narrow peaks.



104

104

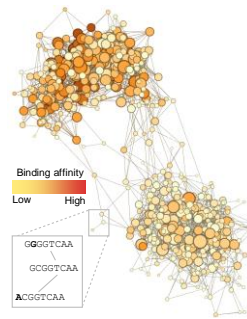
Global peak size correlates with binding site diversity.



Aguilar, Payne, & Wagner, *Nat Ecol. Evol.*, 2017

105

Adaptive landscapes and network science

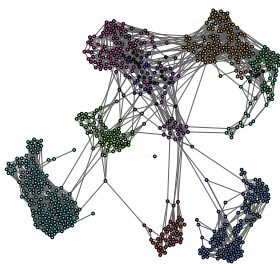


intra-network

- characteristic path length
- diameter
- community structure
- degree assortativity
- ...

106

Adaptive landscapes and network science

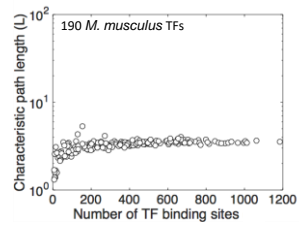


inter-network

- overlap
- interface
- phenotype space covering

107

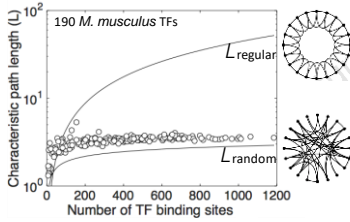
Genotype networks of TF binding sites are "small-world"



Watts & Strogatz, *Nature*, 1998  
Aguilar-Rodríguez et al., *Evolution* 2018

108

Genotype networks of TF binding sites are “small-world”



Watts & Strogatz, *Nature*, 1998  
Aguilar-Rodríguez et al., *Evolution* 2018

109

Summary

Empirical adaptive landscapes of RNA and proteins are somewhat rugged but not prohibitively so

Regulatory landscapes are, with some exceptions, smoother

Our knowledge of landscape topography may be highly biased towards landscapes on which adaptive evolution is possible towards regions of a biomolecule that affect fitness strongly

111

Some further reading

Bank, C. (2022) Epistasis and Adaptation on Fitness Landscapes arXiv preprint arXiv:2204.13321, arxiv.org

Szendro, I. G., Schenk, M. F., Franke, J., Krug, J., & De Visser, J. A. G. (2013). Quantitative analyses of empirical fitness landscapes. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(01), P01005

Obolski, U., Ram, Y., & Hadany, L. (2017). Key issues review: evolution on rugged adaptive landscapes. *Reports on Progress in Physics*, 81(1), 012602.

Kinney JB, McCandlish DM. (2019) Massively parallel assays and quantitative sequence–function relationships. *Annu Rev Genomics Hum Genet*.20:99–127.

113

A successful CSSS project



José Aguilar-Rodríguez Massimo Stella Leto Peel

The architecture of an empirical genotype-phenotype map

José Aguilar-Rodríguez, Leto Peel, Massimo Stella, Andreas Wagner, and Joshua L. Payne (2018), *Evolution* 72: 1242–1260.

110

Outlook

Adaptive «seascapes»

Extradimensional by-passes

Co-evolutionary landscapes

Landscape mapping versus evolution

Landscapes of engineering problems

112