Genotype/Phenotype Modelling of Evolutionary Landscapes in Spatial Patterning

Bhavin Khatri[$], Richard Sear[$]

and Tom McLeish[*],

[$]*Dept. of Physics, University of Surrey.*

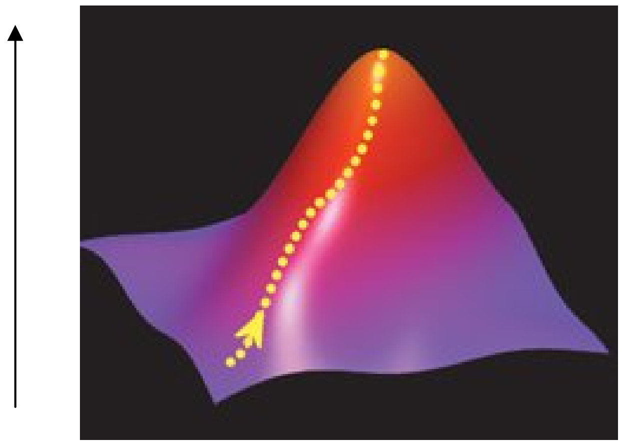
[*]*SMPC, Departments of Physics and Chemistry and BSI, Durham University*

Durham University

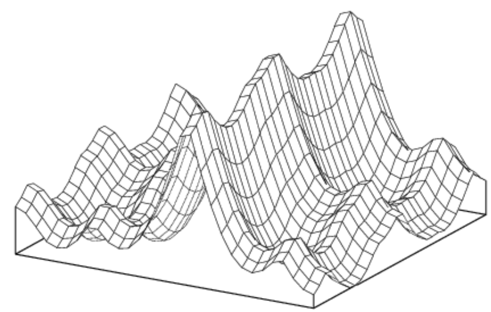
Phenomenological Evolutionary Landscapes

"Mount Fuji"

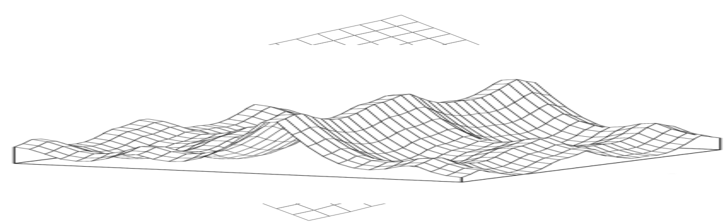
Fitness



Rugged (NK, Spin-glass)



Neutral



Scale?

Questions

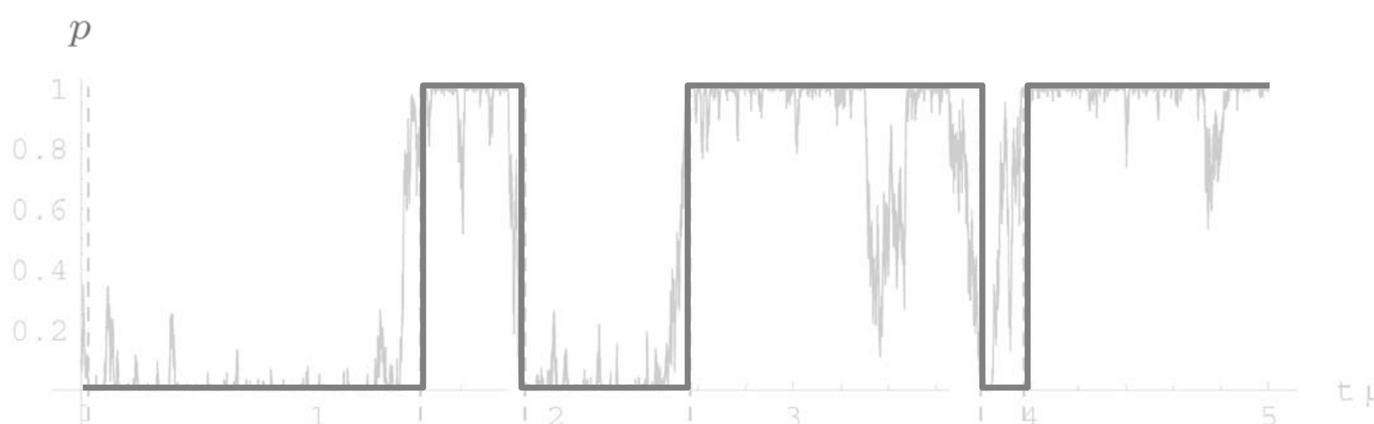
- What do realistic fitness landscapes look like?
  - What is the importance of mapping from sequence to function?
- What is the important scale of fitness?
- Is there an underpinning structure to convergence?
  - Link to ergodicity

Evolution for finite populations

- For μN<<1 evolution proceeds by sequential fixation of rarely occurring mutations, through combination of selection and randomness in birth & deaths (*genetic drift*)

All *A*

All *a*



g, *BMC Bioinformatics* 2007,
ppl 6):S7

= 0.05; $N\delta F$ = 0.5

- Probability of fixation of mutant with fitness difference δF relative to wildtype

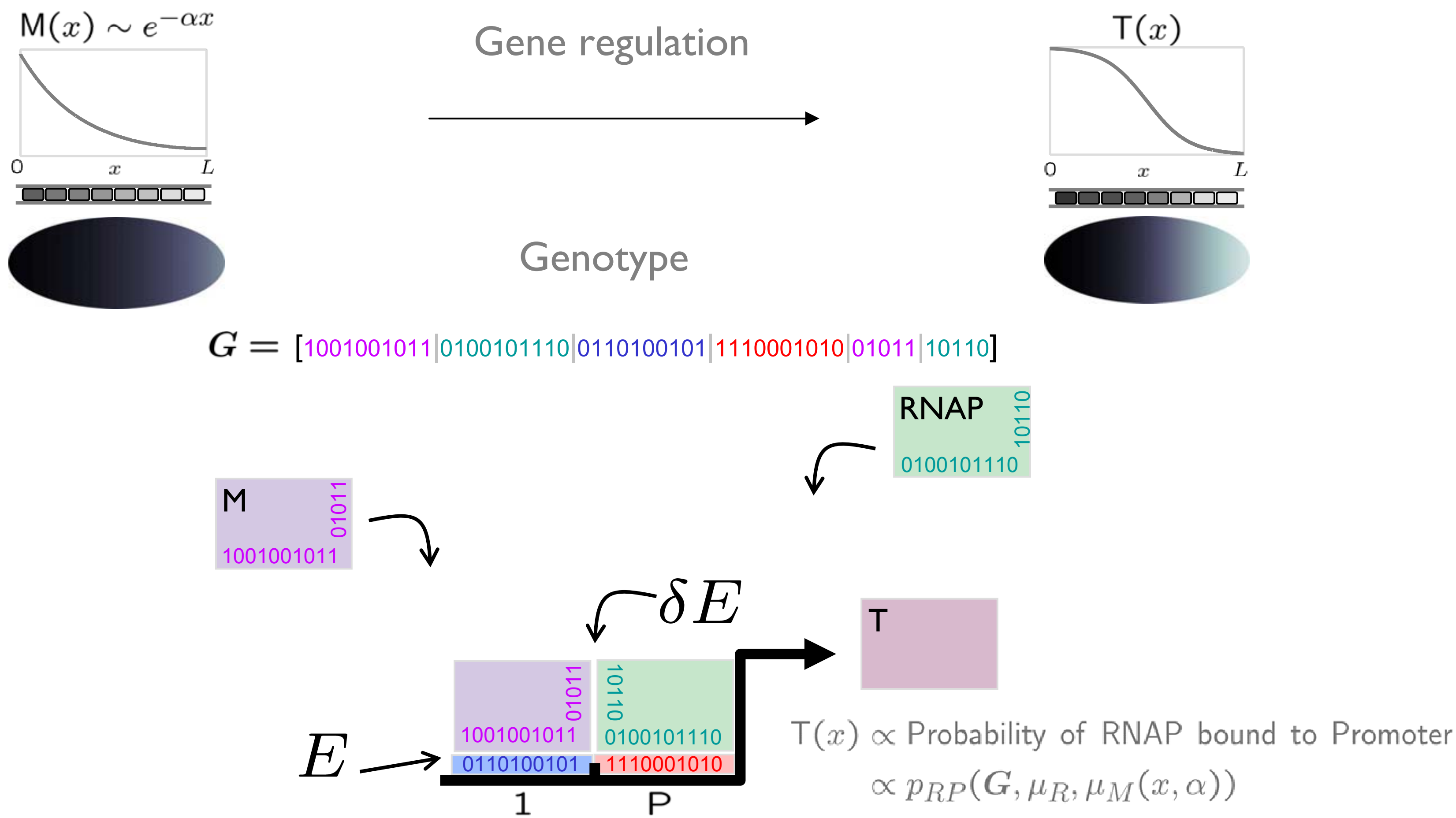
Kimura, M.
On the probability of fixation of mutant genes in a population.
Genetics, **1962**, 47, 713-719

$$\varphi(\delta F) = \frac{1 - e^{-2\delta F}}{1 - e^{-2N\delta F}}$$

Monte Carlo simulations

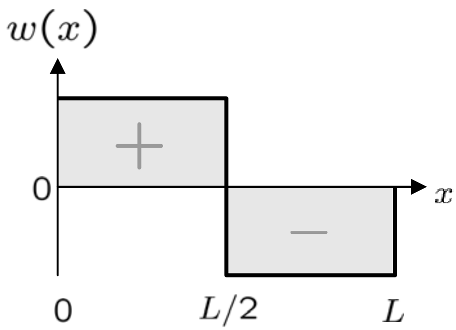
Biophysically motivated model of genotype-phenotype map

- Pattern anterior of cellularised embryo



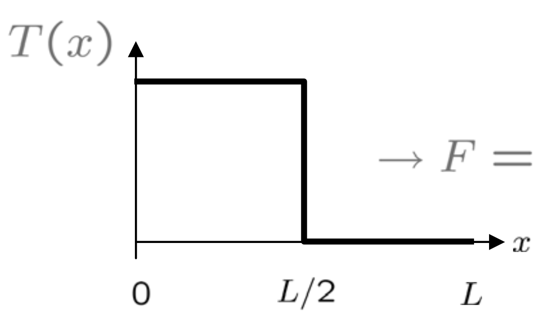
Fitness function of concentration profile

- Choose functional that selects for contrast



$$F = \mathcal{F}\left[T(x)\right] = \frac{1}{\frac{L}{2}\,\mathrm{max}_x\{T(x)\}}\left(\int_0^{L/2} T(x)\mathrm{d}x - \int_{L/2}^{L} T(x)\mathrm{d}x\right)$$

e.g.



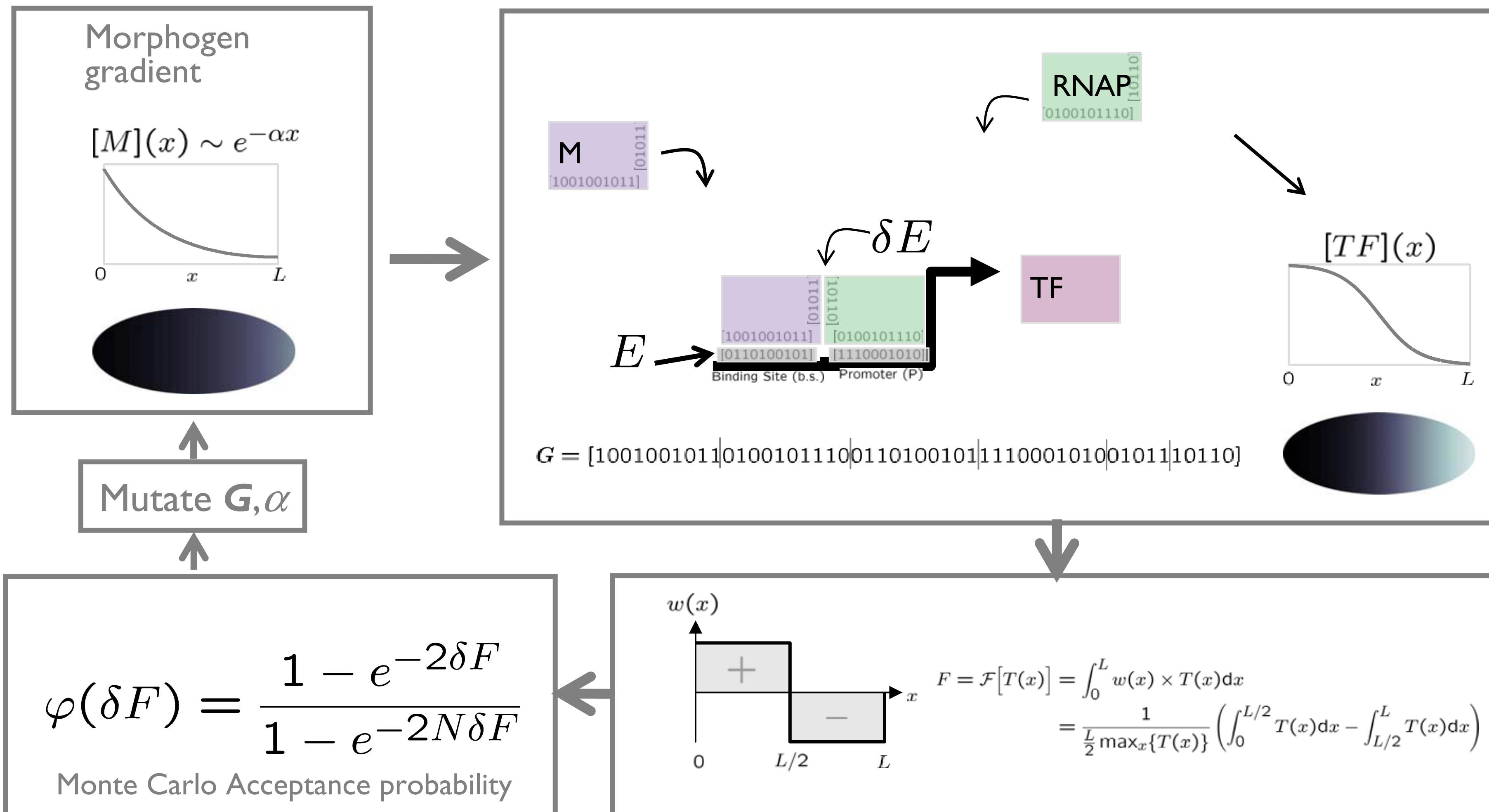
$\rightarrow F = 1$

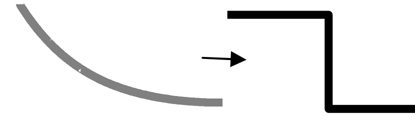
Functional does not change with time
$\rightarrow$ Evolution in a constant environment

*

B.S. Khatri, T.C.B. McLeish and R.P. Sear, "Statistical mechanics of convergent evolution in spatial patterning", *PNAS*, **106**, 9564-9569 (2009)

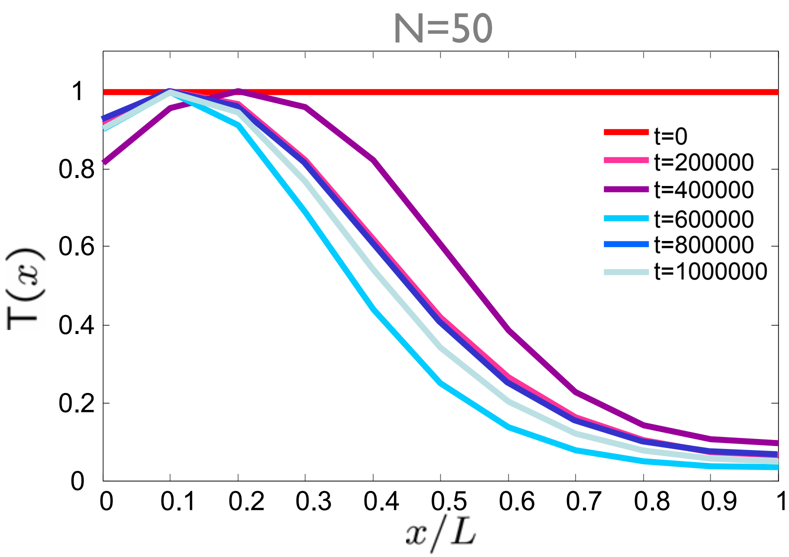
Overview of Model

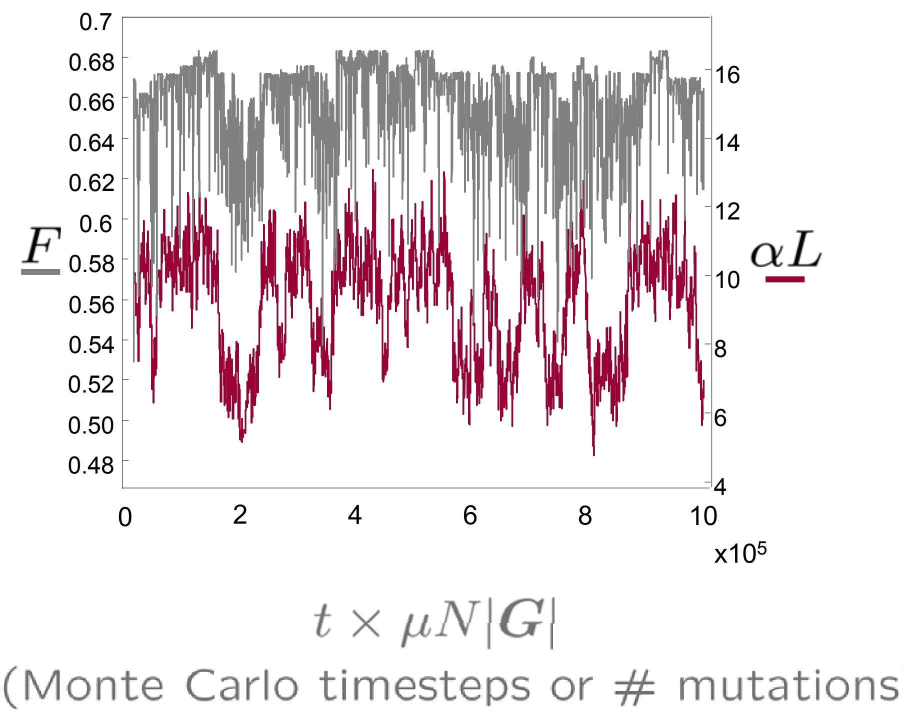




Results:
Simulations with anterior patterning functional

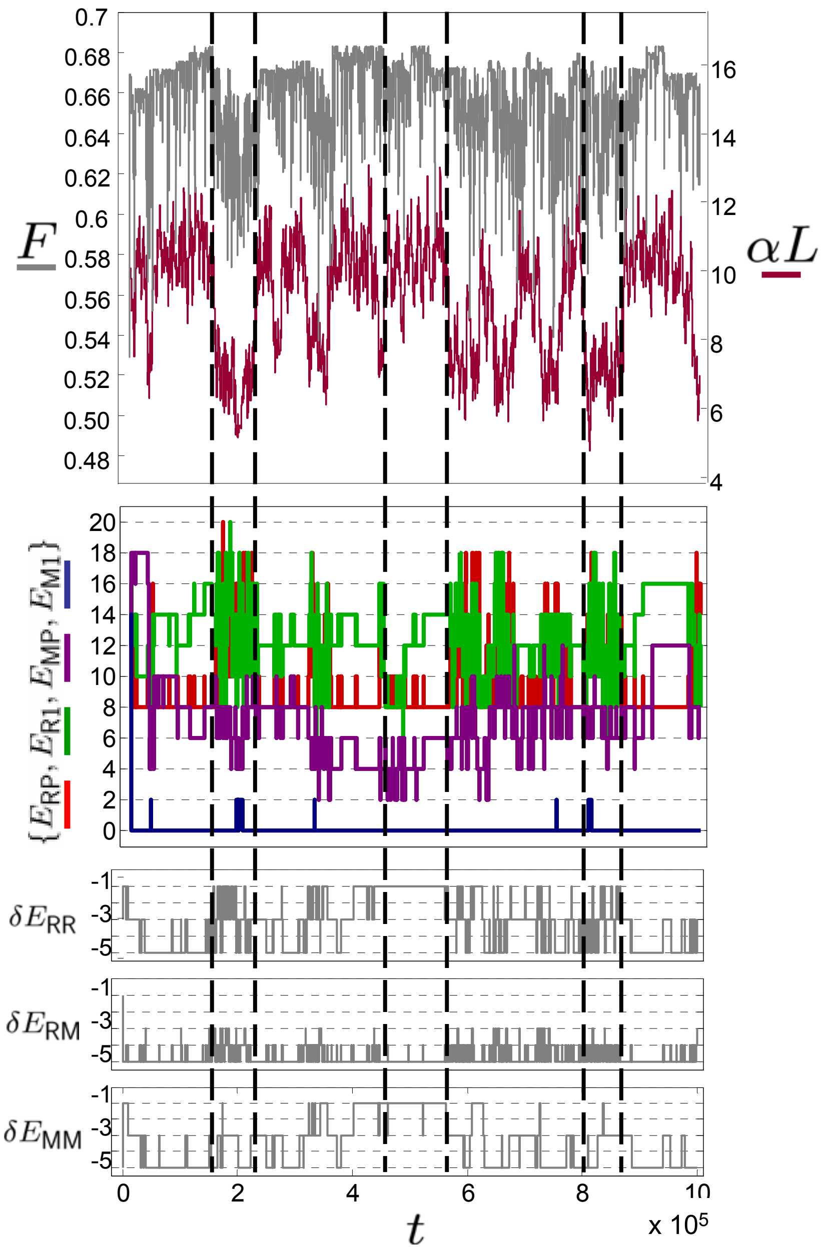
- Each run starts with random genome (50 binary bps $\Rightarrow 2^{50}$ points in genotype space) with $10^{\wedge 7}$ attempted mutations & $\alpha_0=10$
- Each time step a mutation in ***G*** or α - not both - α mutated continuously



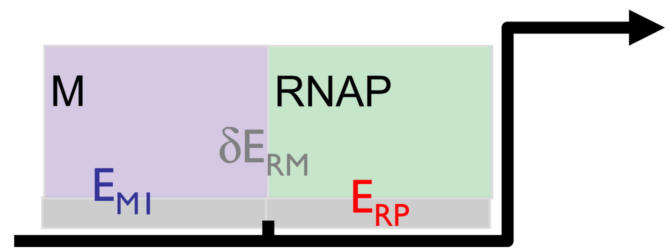


- Bistability
- Two emergent 'preferred' solutions ($\alpha \approx 7$ & $\alpha \approx 10$)
- $F(\alpha \approx 7) < F(\alpha \approx 10)$

Results: Anterior Patterning



- Single global phenotype: threshold mechanism of co-operative binding of M-RNAP to P



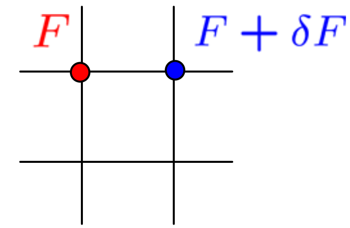
- Critical & non critical E's & δE's →larger variation in non-critical
- $\alpha \approx 7$ solution substitution rate higher than $\alpha \approx 10$ indicates difference in local curvature and/or roughness

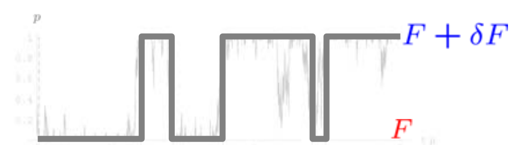
Statistical Mechanics of the Evolution of Finite Populations

- In *equilibrium*, assuming microscopic reversibility at genotype level, detailed balance is obeyed

$$p(F)\varphi(\delta F) = p(F+\delta F)\varphi(-\delta F)$$

$$\Rightarrow \frac{\varphi(\delta F)}{\varphi(-\delta F)} = \frac{p(F+\delta F)}{p(F)} = e^{\nu \delta F}$$



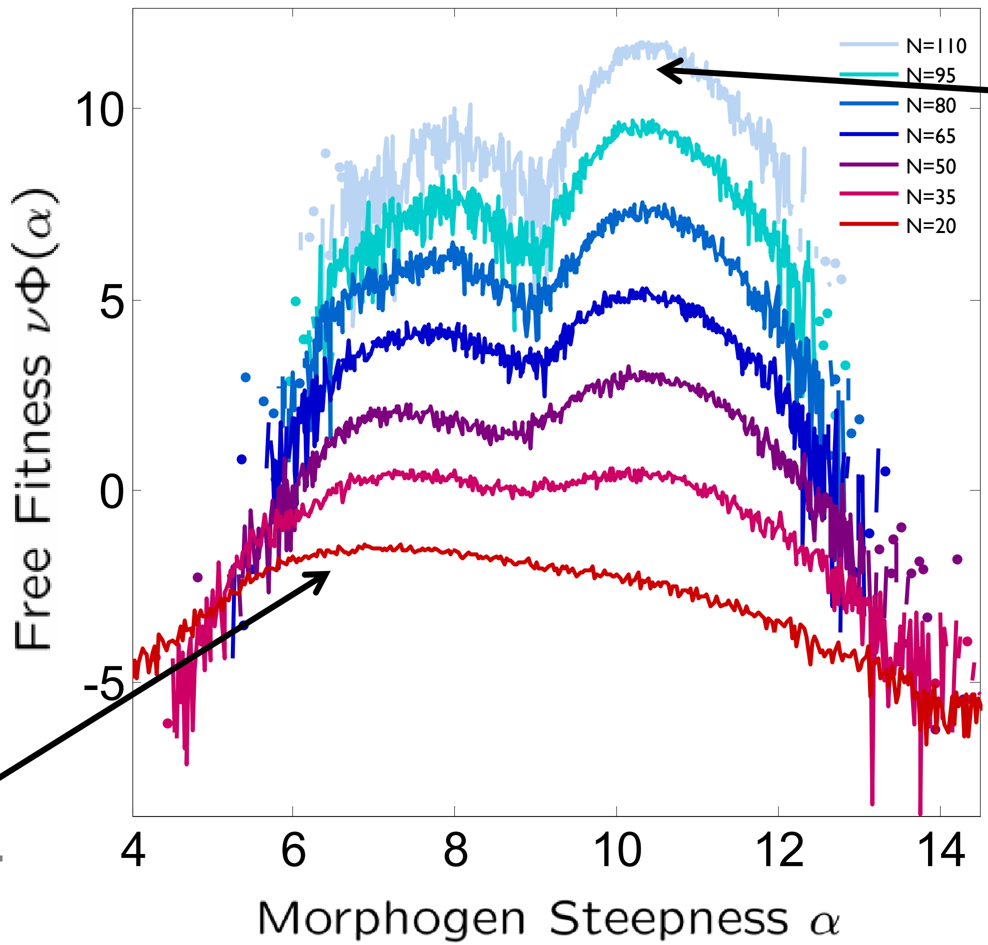


- Equilibrium distribution is Boltzmann : $`k_B T` \to \nu^{-1} \approx 1/2N$
- Differences in fitness $\delta F << 1/N$ are neutral
- Energy function: $\Phi = F + S/\nu$ (*Free Fitness*)

Sella & Hirsh, (2005), PNAS, **102**, 9541–9546.

Free Fitness Landscape $\Phi(\alpha)$

$$\Phi = F + \frac{1}{\nu} S = \frac{1}{\nu} \ln Z \sim \frac{1}{\nu} \ln p$$

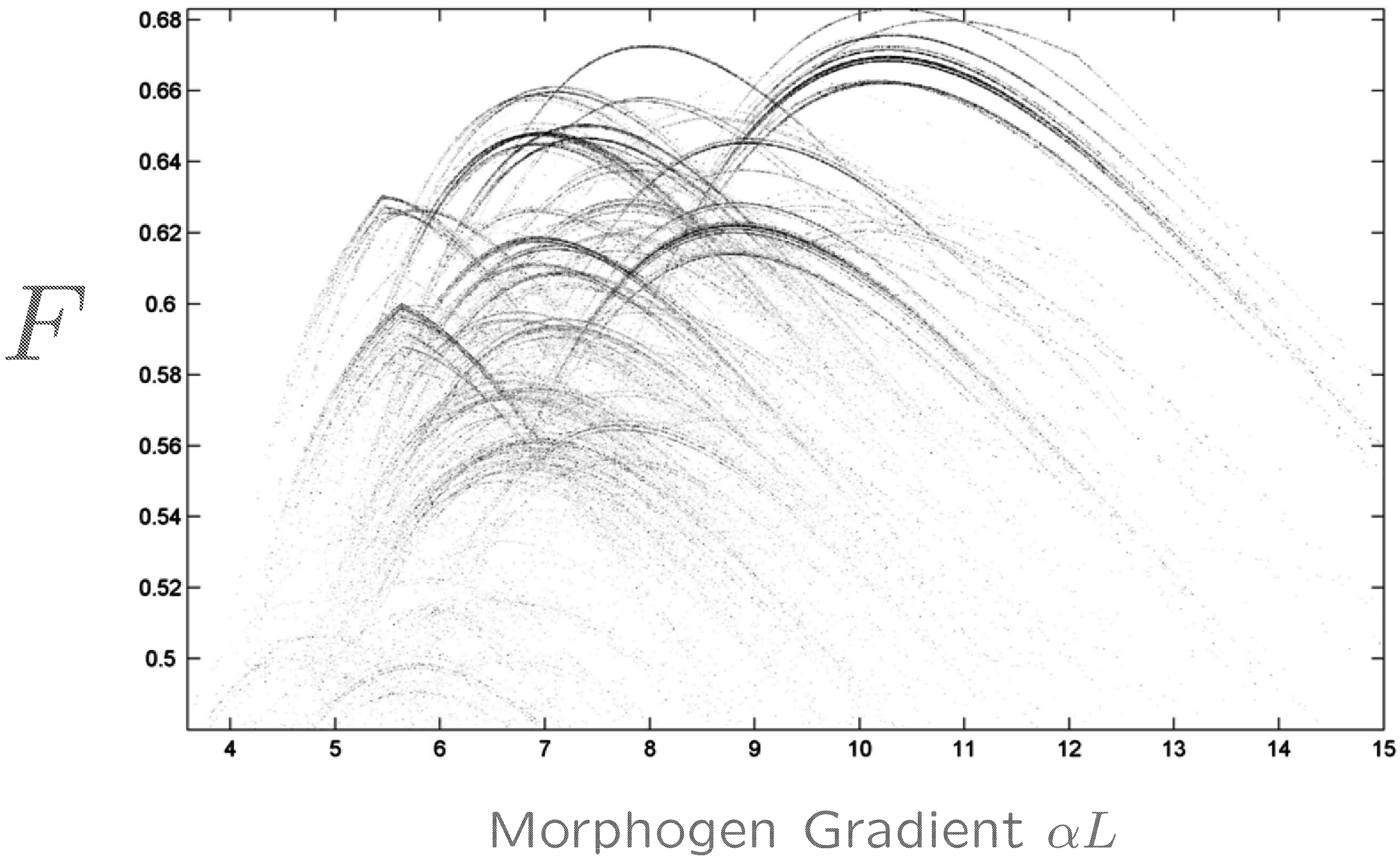


High *N*, Fitness dominates: Convergence to high fitness phenotype

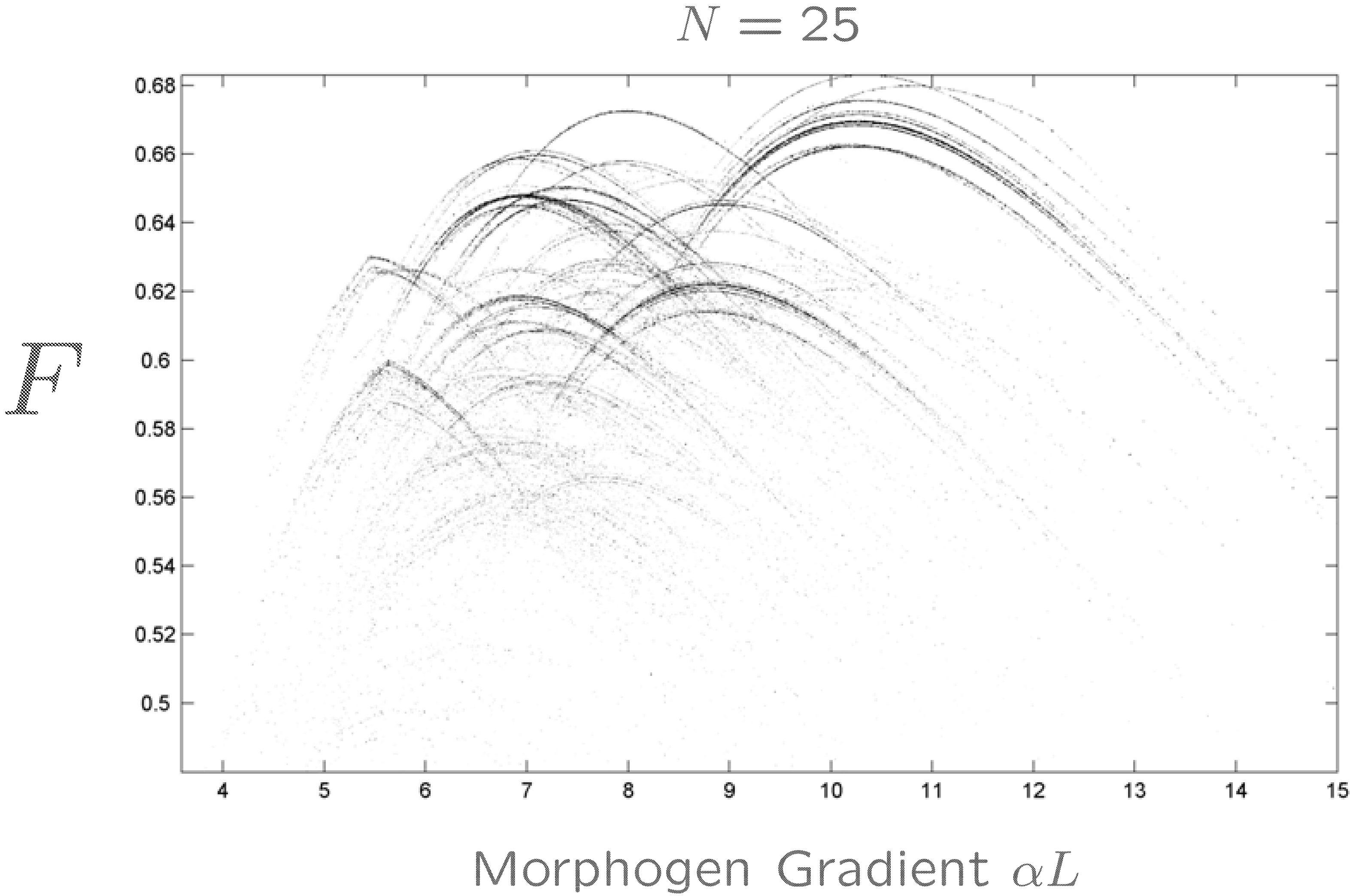
Low *N*, Entropy dominates: Convergence to **sub-optimal** phenotype

The Fitness Landscape

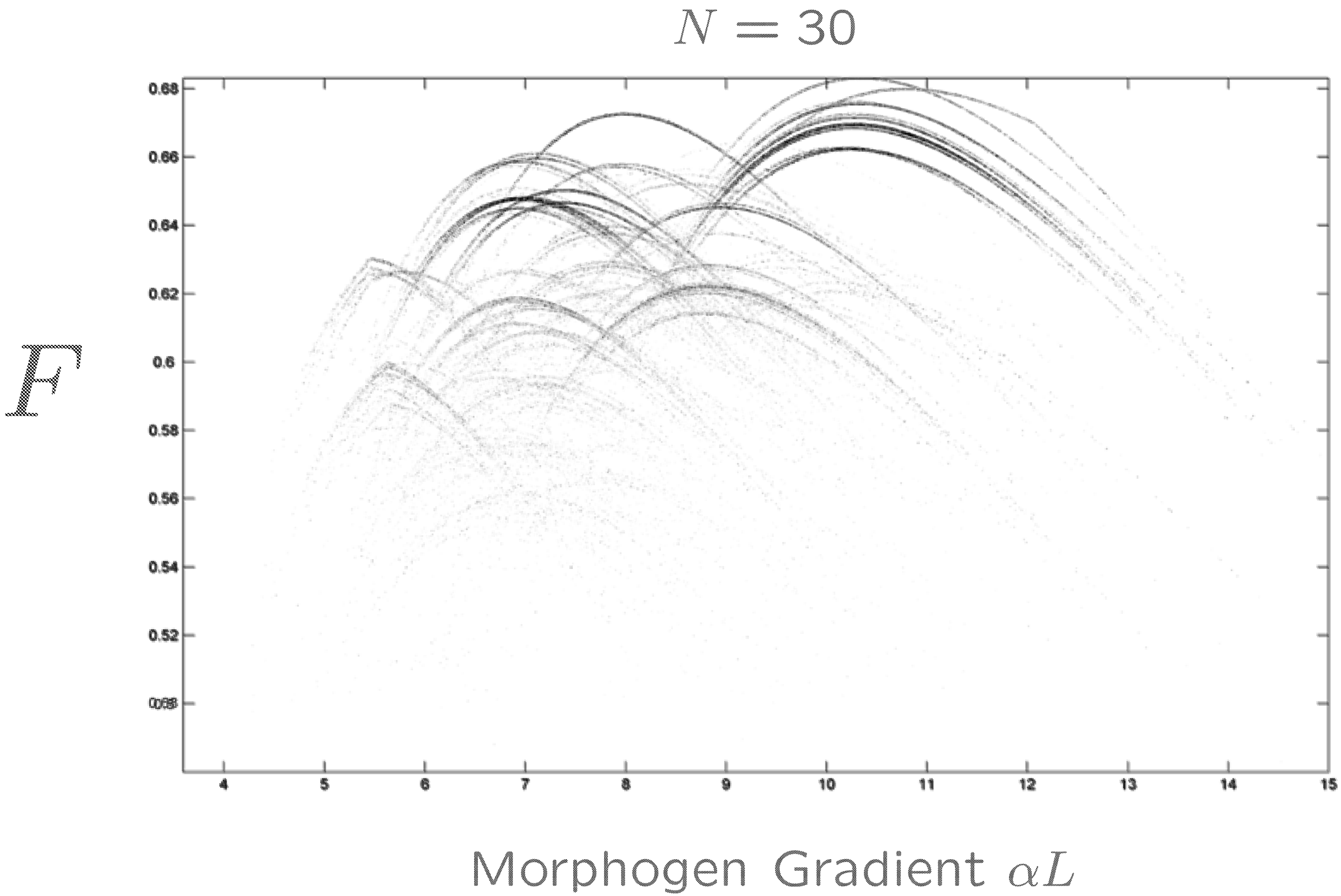
$N = 20$



The Fitness Landscape



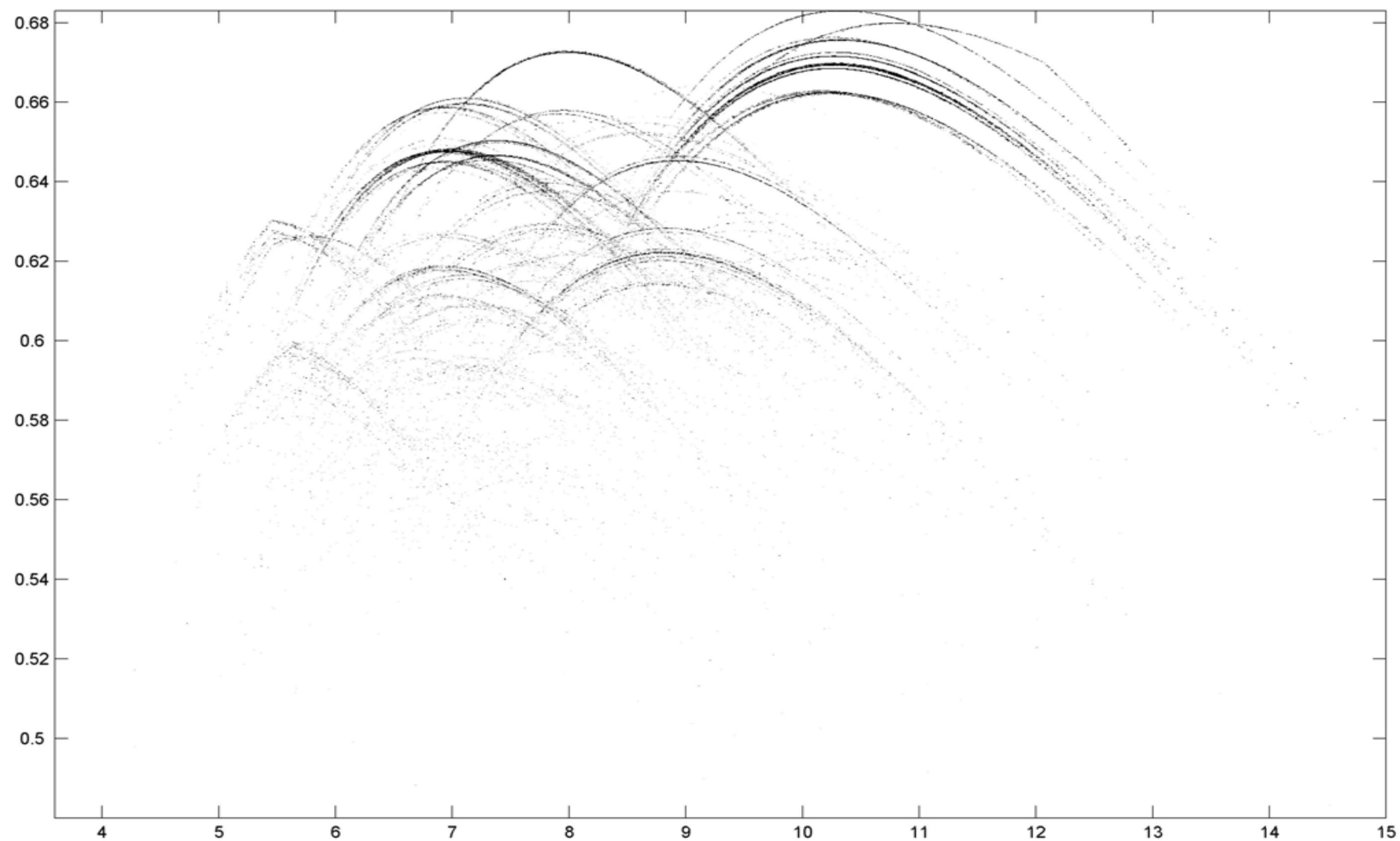
The Fitness Landscape



The Fitness Landscape

$N = 35$

F

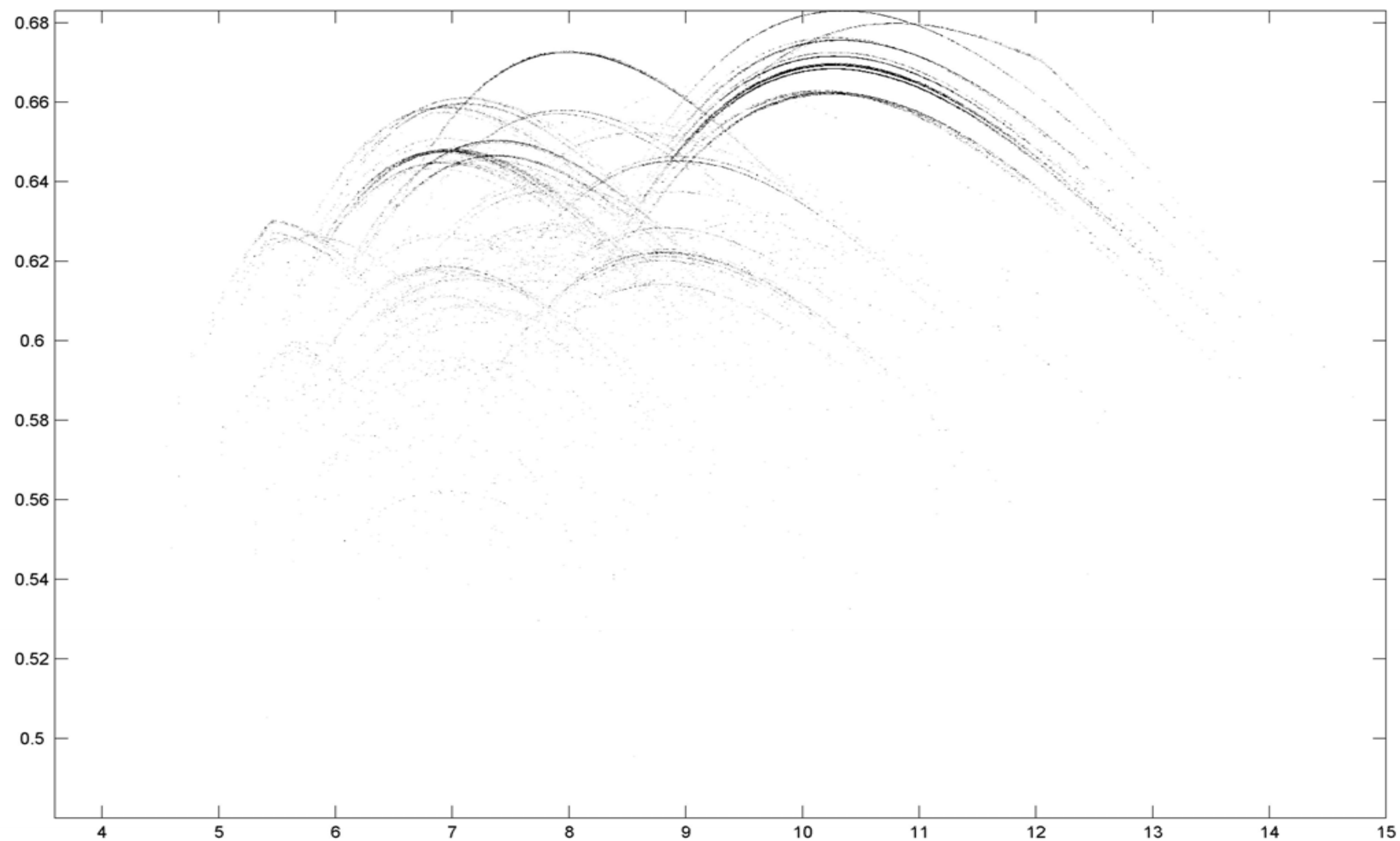


Morphogen Gradient αL

The Fitness Landscape

$N = 40$

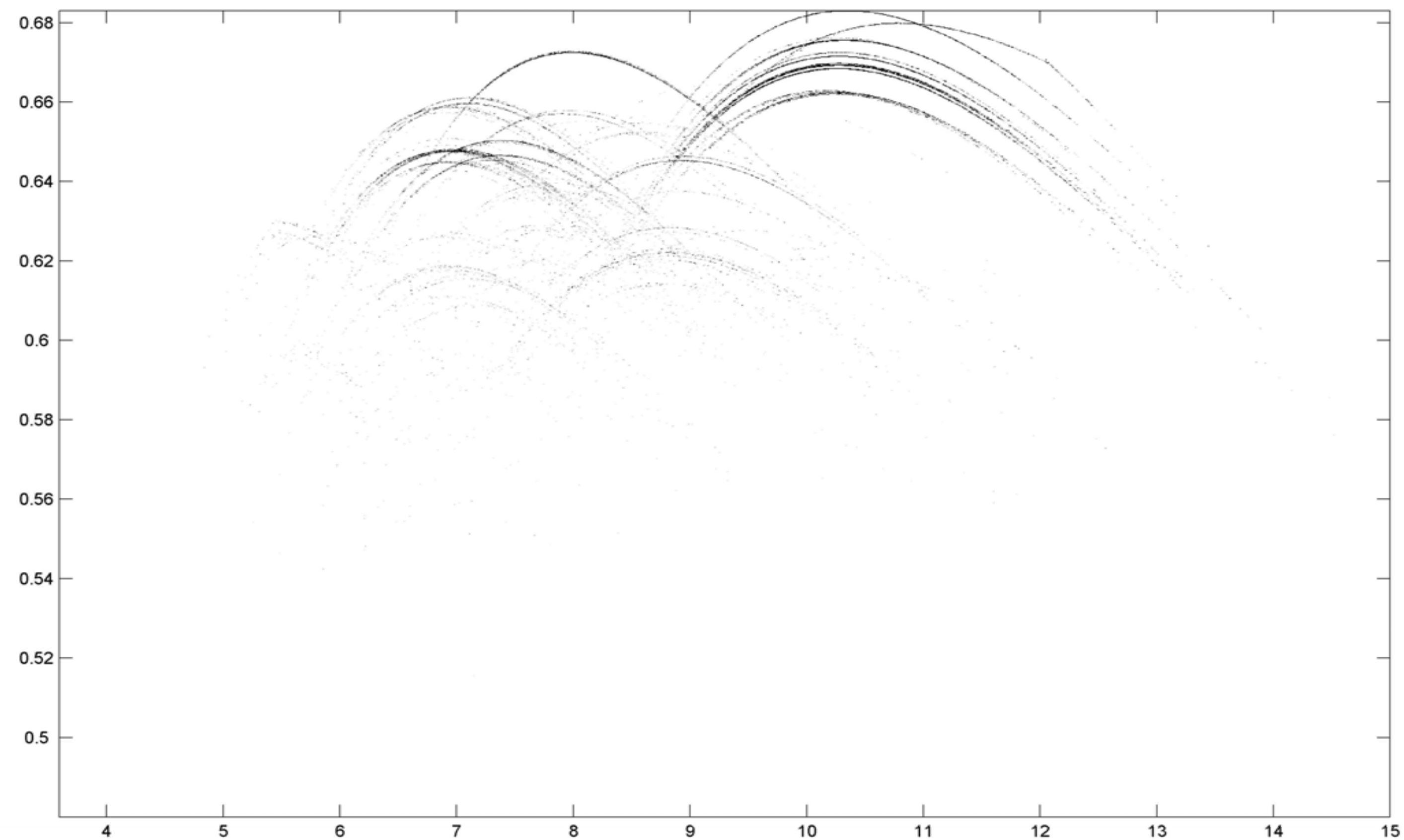
F



Morphogen Gradient αL

The Fitness Landscape

$N = 45$

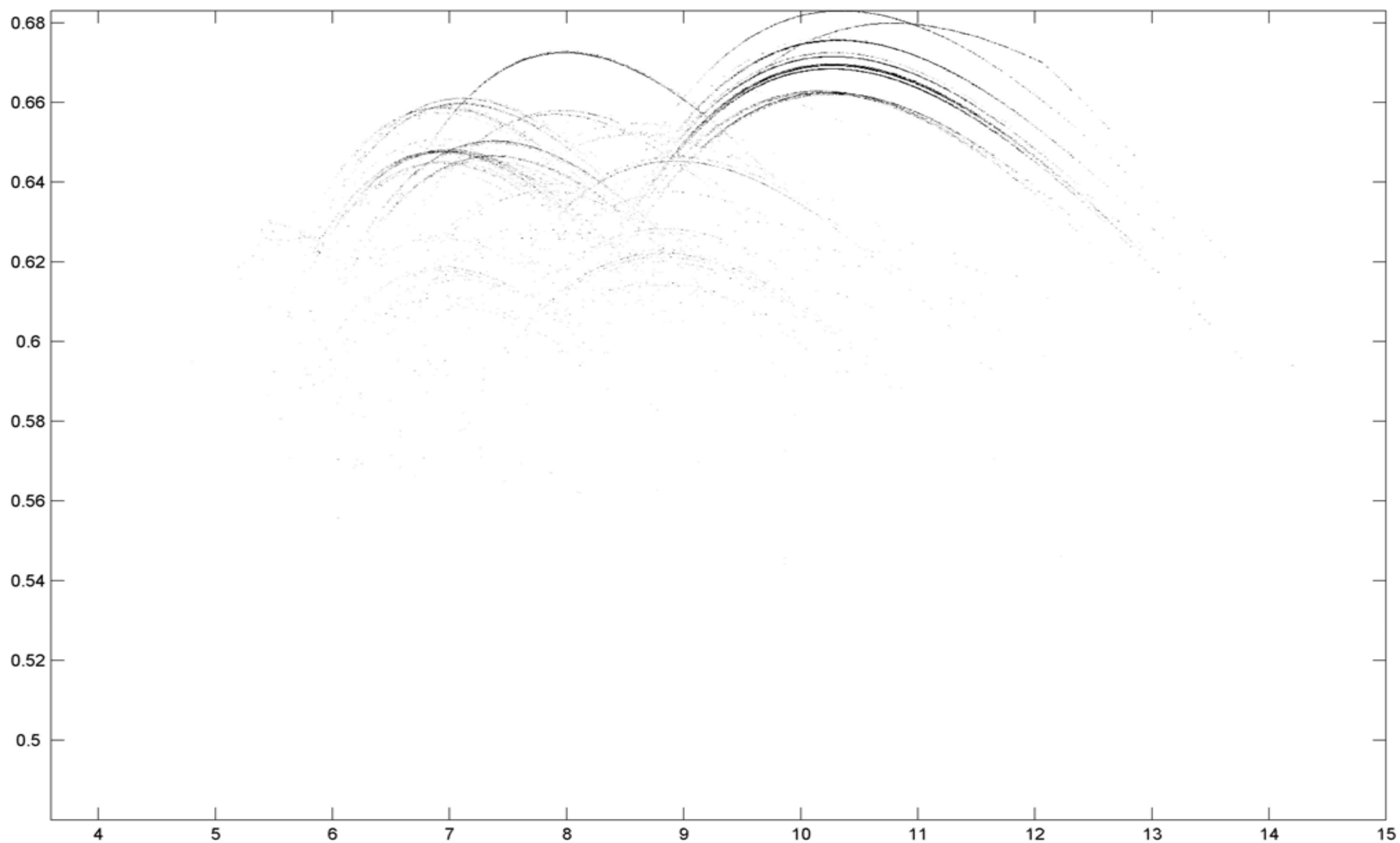


F

Morphogen Gradient αL

The Fitness Landscape

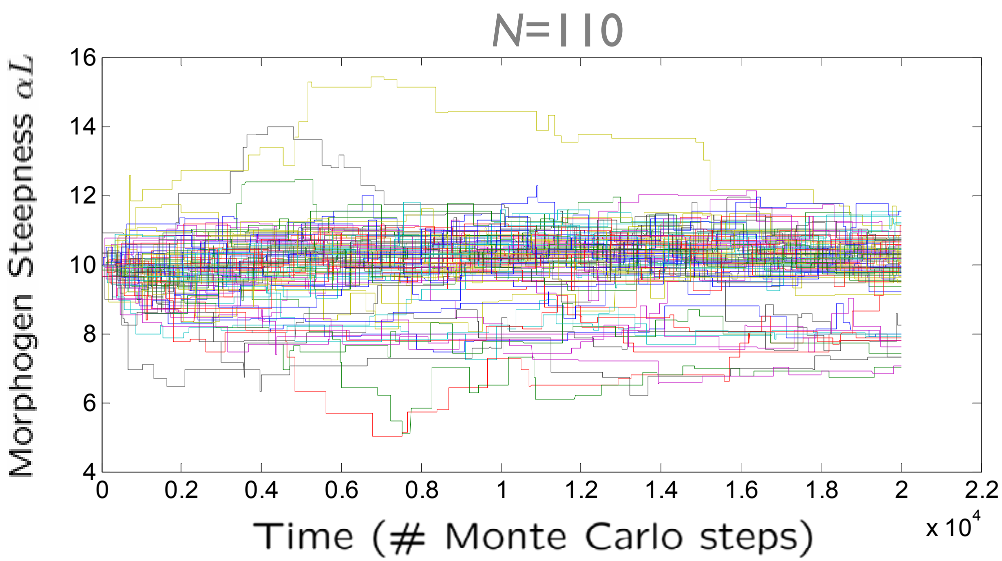
$N = 50$

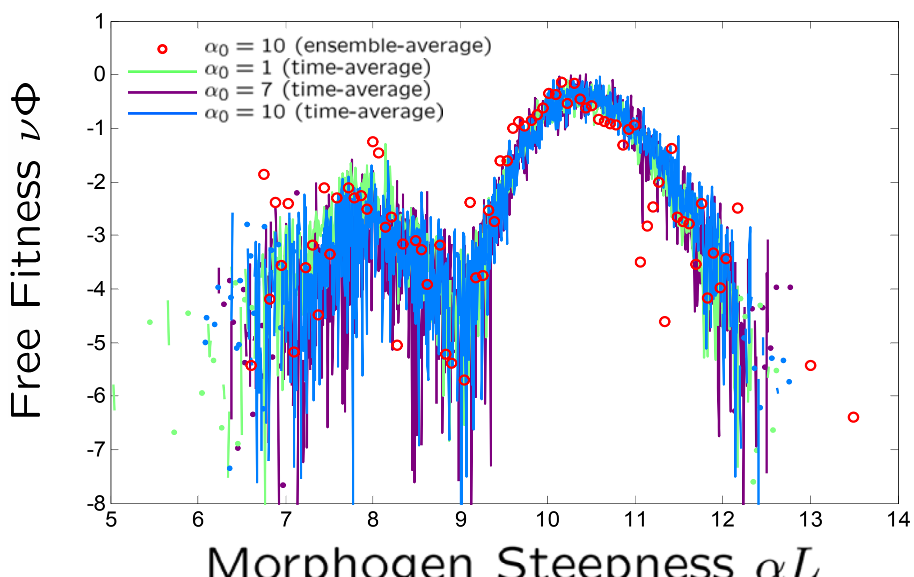


F

Morphogen Gradient αL

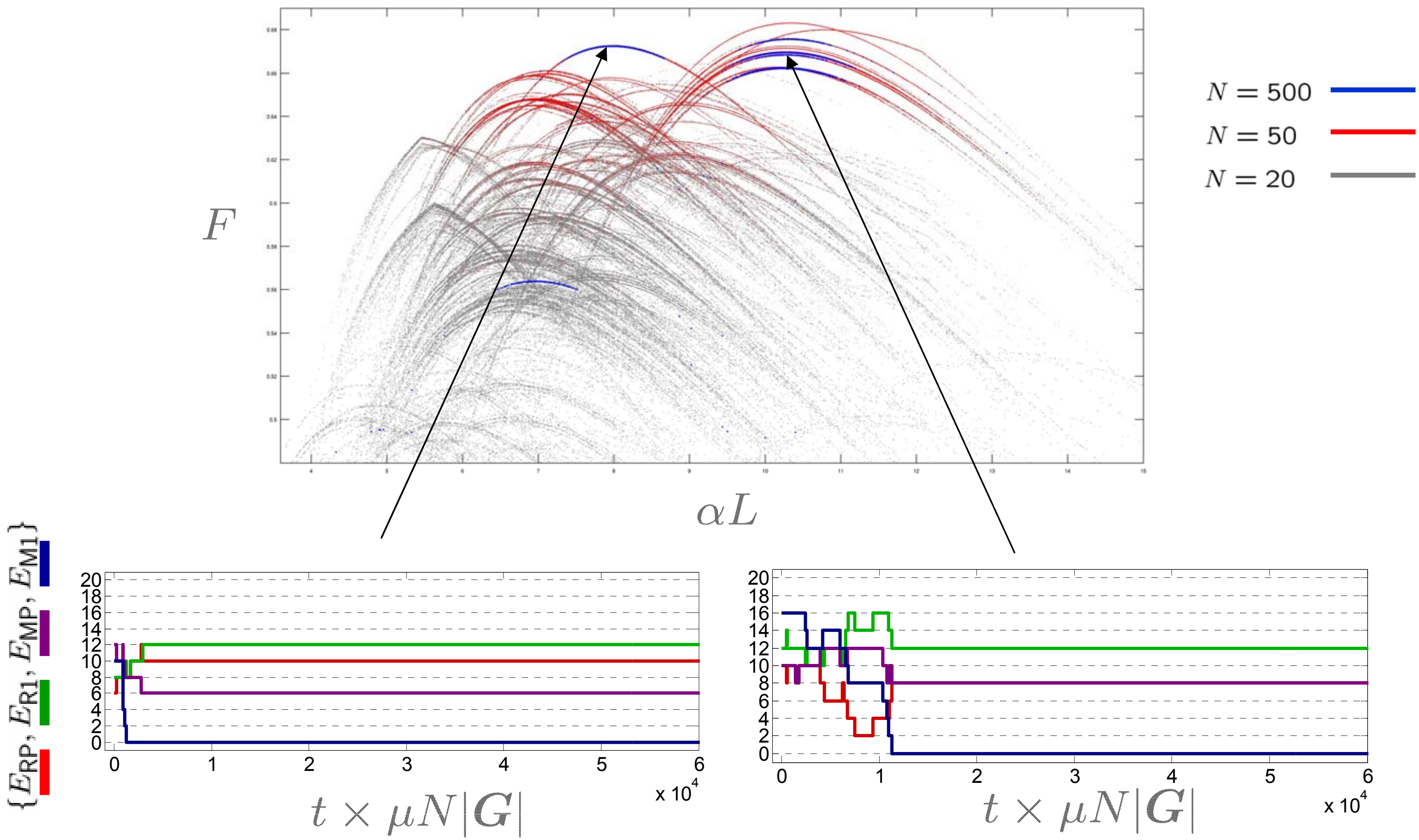
Ergodicity





- At low population sizes simulations are ergodic (N<~100)
- 10^7 mutations can only explore a very small fraction of phase space (2^50)
- Underlying symmetry of genotype space??

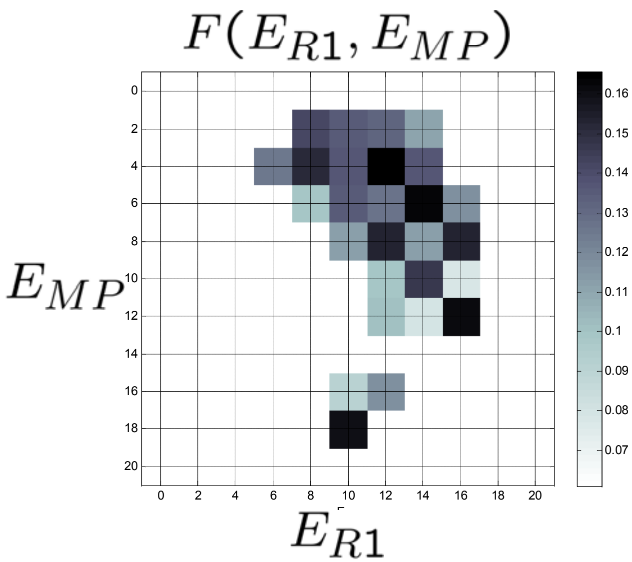
Quenched disorder at High N



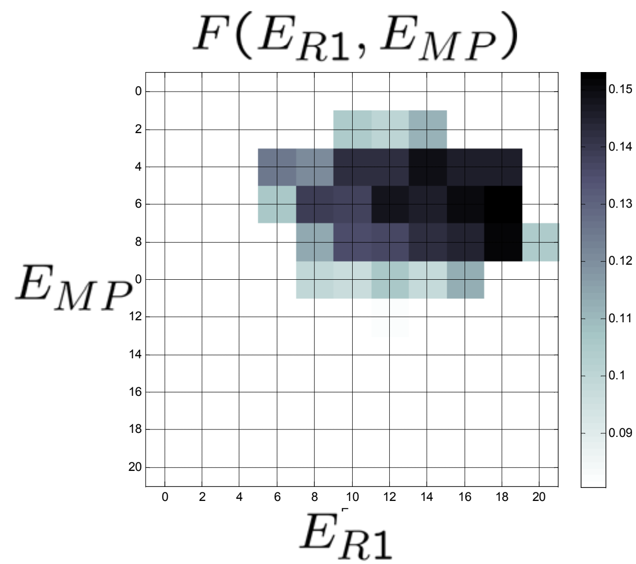
Non-critical energies are quenched

Landscape locally rough

N=50



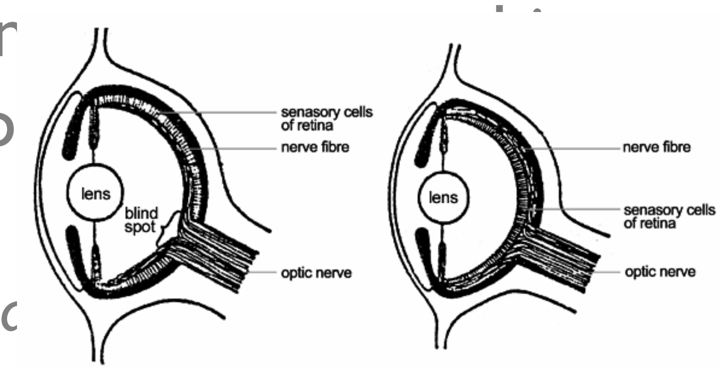
High α (Rough)



Low α (Smooth)

Conclusions

- Emergent complexity even for minimal genotype-phenotype map of gene regulation
  - *difficult to predict a priori 2 preferred morphogen gradients*
- Entropy from the genotype-pher
  evolution towards sub-optimal p
  population sizes;
  - *Are most organisms sub-optimally adc*



- Quenched disorder of non-essential binding energies at high *N*
  - *Conservation does not necessarily imply function*
  - *More conserved phenotypes for large populations?*

Acknowledgments

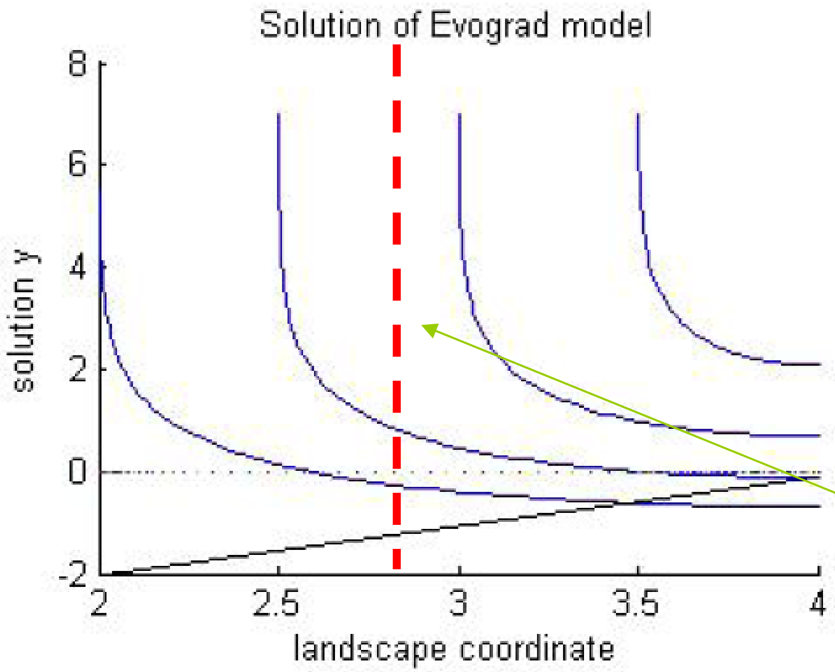
- The John Templeton Foundation's, Cambridge Templeton Consortium

Future Work

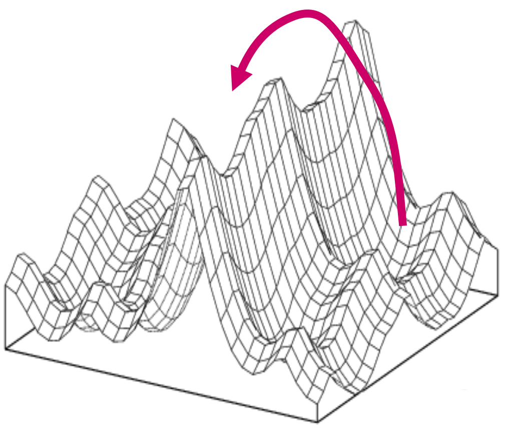
- Expect more complicated and realistic gene regulatory networks may have large and non-trivial entropic contributions to free fitness
- As combinatorial complexity increases will glassy nature of landscape disappear or get worse?
- Useful methodology for probing questions of robustness and evolvability, where robustness~entropy

Evolutionary Bottlenecks

$$\frac{dx(t)}{dt} = \frac{\partial F}{\partial x} N e^{-\Delta N}$$



Rugged (NK, Spin-glass)

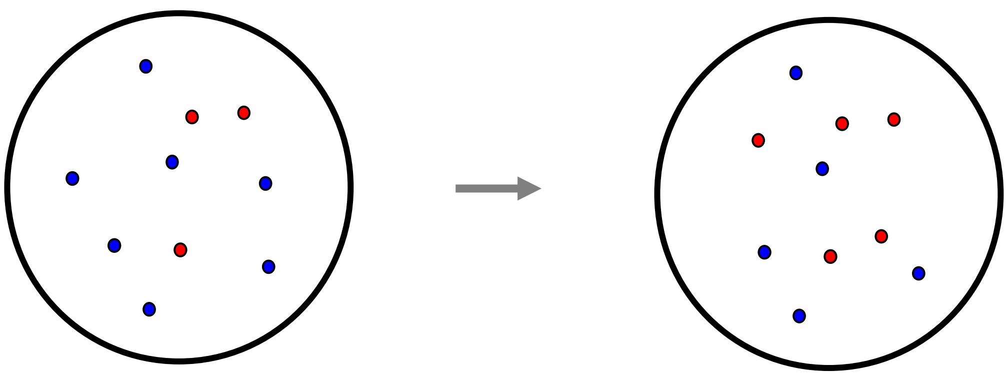


Boundary of survival

Statistical Mechanics for Populations of Finite Size

- Stochastic fluctuations in gene frequencies due to randomness in reproductive success become important at low population sizes
- **Example:** Two neutral alleles A and a with frequencies *p* and 1-*p* in a population of *N* individuals with asexual reproduction

Randomly sample N alleles



$$\langle p' \rangle = p$$

$$\langle (p' - \langle p' \rangle)^2 \rangle = \frac{p(1-p)}{N}$$

Gene-frequencies Binomially distributed

Generation *n*
f(A)=*p*
f(a)=1-*p*

Generation *n*+1
f(A)=*p*'
f(a)=1-*p*'

Wright-Fisher Process

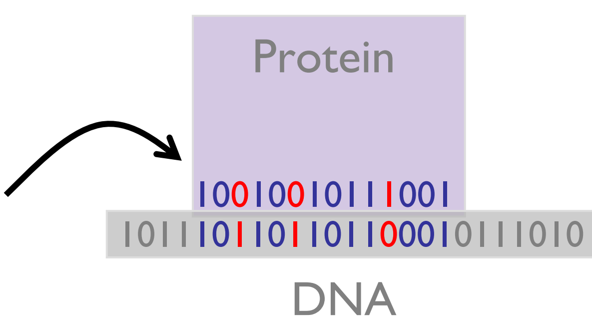
Assumptions

- $\mu N \ll 1$, but $N \gtrsim 10$
- Asexual Haploid Population
- Only point mutations (no indels, gene duplications or recombination)
- Cellular concentrations are large to avoid intrinsic noise
- Constant environment (Equilibrium)
- No competition with other species
- No spatial or geographic variation

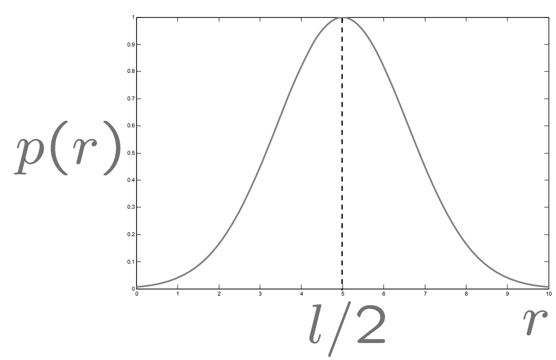
Genotype to Phenotype Map

- Phenotype: any function of sequence
- Selection acts on phenotype
- Many to One: e.g. protein binding DNA

$$E(r) \sim \epsilon r$$



$$p(E) \sim p(r) = \frac{1}{2^l}\binom{l}{r}$$



2 Key Ingredients?

Map from Sequence to Function
(Genotype-Phenotype Map)

Effective Statistical Mechanics
due to finite population sizes
(Sella & Hirsh, PNAS, 2005)

Emergent Structure & Phenomenon
Analogous to Condensed Matter Physics

Entropy

Ergodicity

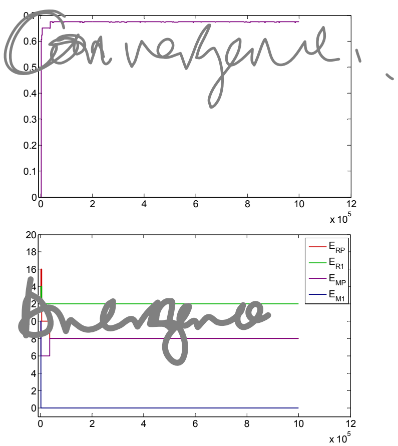
Glass-like
evolution

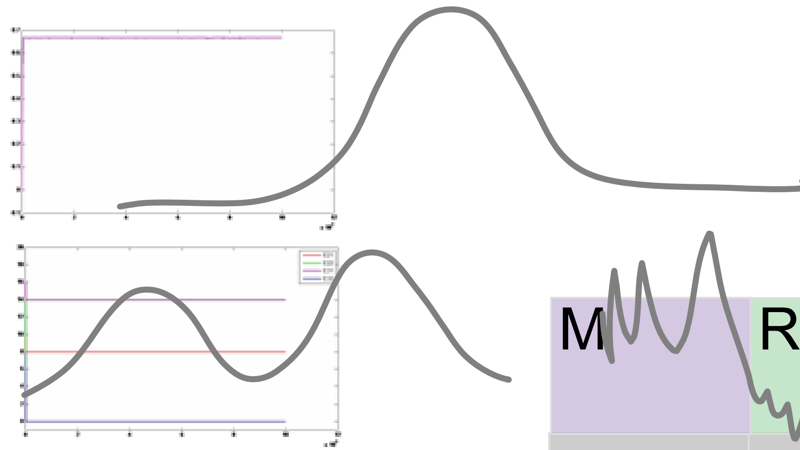
What is Evolution?

- **Life:** Reproducing organism carrying information, e.g. in DNA, about how to survive in an environment
- **Selection:** those organisms that have the best information about surviving reproduce best/fastest (*survival of the fittest*)
- **Mutation:** random changes in information
- *Mutation + Selection* $\rightarrow$ adaptation to best information

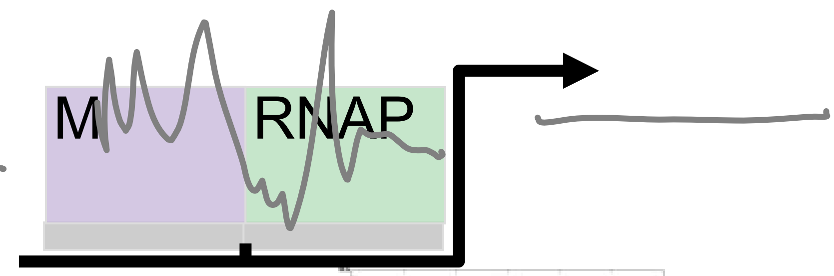
Very high N

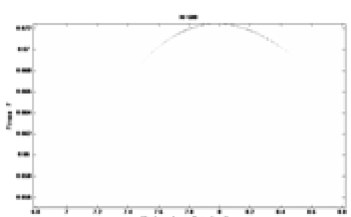
$N = 500, \delta t = 20\text{kGn}$

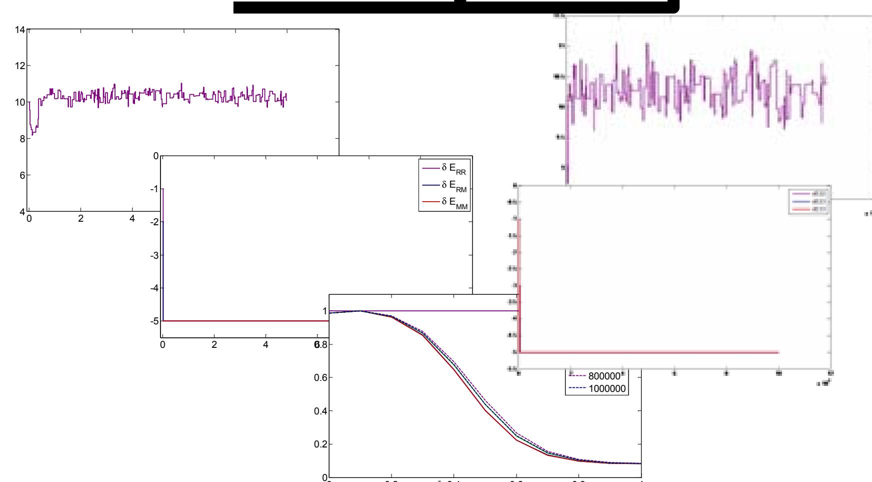




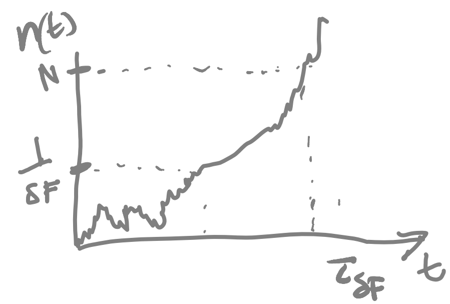
$t \times \mu N|\boldsymbol{G}|$ $t \times \mu N|\boldsymbol{G}|$







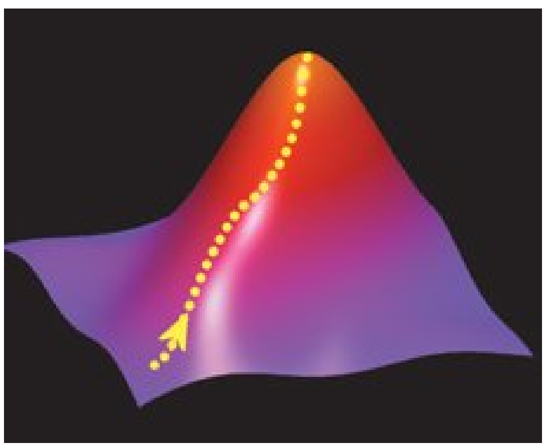
x/L


n(t)
N
1
SF
t

Statistical Mechanics of Finite Population-size Fluctuations

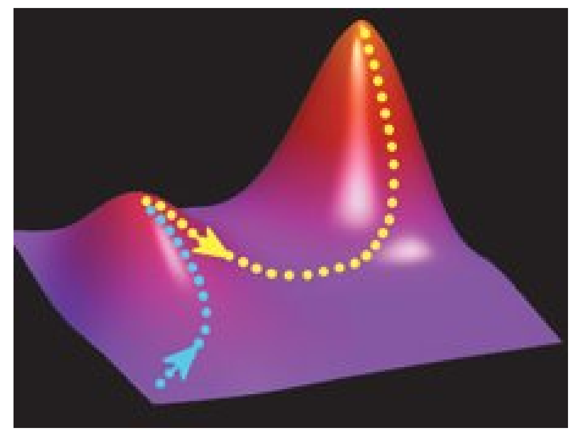
Uhill adaptation

Fitness



Large population size

Up & Downhill steps



Small population size