

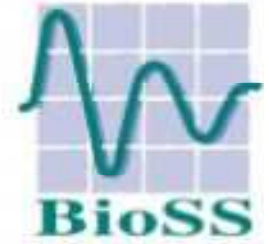
# An analytical method for neutral evolution in a type space

**Daniel Lawson**

Bioinformatics and Statistics Scotland

*Work with*

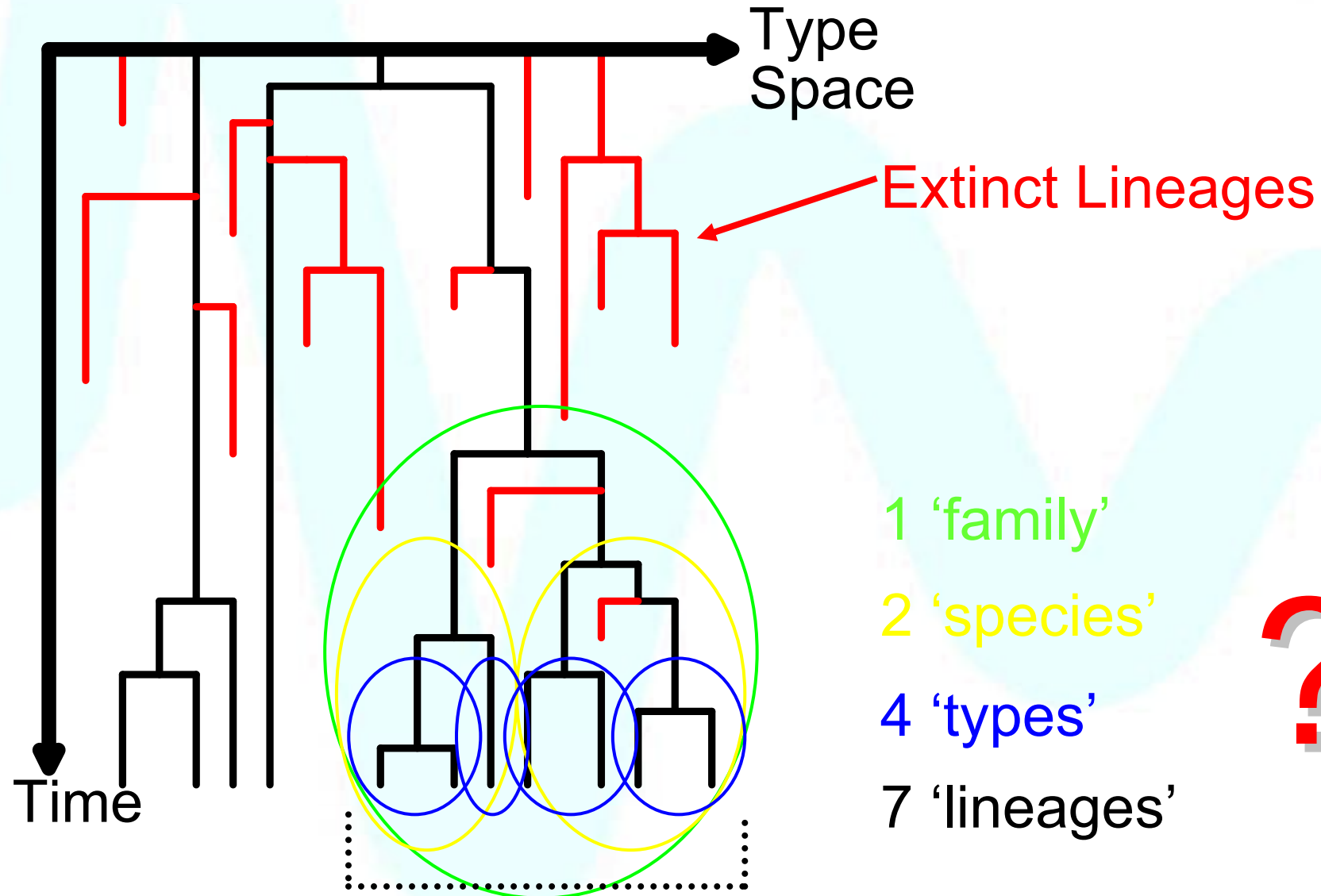
**Henrik Jeldtoft Jensen,**  
Imperial College London

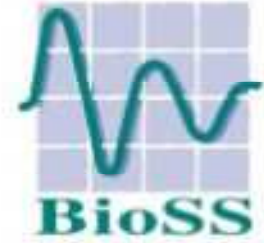


# Background

- *Neutrality* is important in studying the evolution *within a species*
- Because *Species as a concept* doesn't always apply!
- Description instead at the level of *distributions of individuals* are required.
- Our method describes a phenotype.

# The Lineage





# The Neutral Model

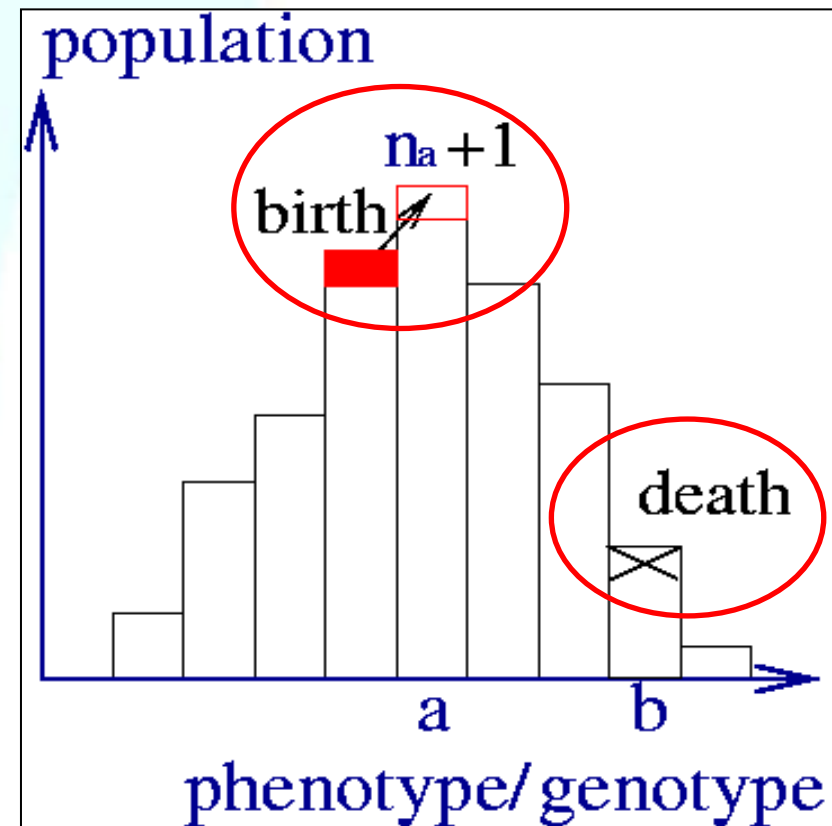
- Assume that all individuals are 'equal'
  - Valid for Phenotypes that do not have function
  - Genotype not coding for protein synthesis (**12% of Human DNA is variable!** *Redon et al. Nature. doi:10.1038/nature05329* )
- Each individual has the same probability to die ( $p_k$ ), or give birth ( $p_b$ ), in a time step
  - For simplicity, assume the total population ( $N$ ) has reached equilibrium ( $p_k = p_b$ )
- Reproduction is asexual and imperfect – small changes can occur

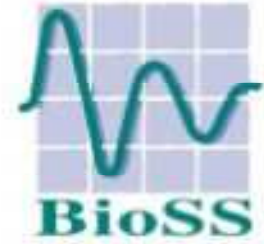
# A neutral evolution model

- Consider  $N$  individuals each labeled by a type position:

## TIMESTEP:

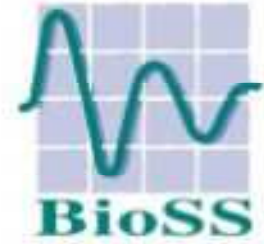
1. Pick an individual (from  $N$ ) and mark it to die.
2. Pick an individual (from  $N$ ) and copy it. With probability  $p_m$  **Mutate to a similar type.**
3. Kill the marked individual.





## The Lineage (2)

- If mutations occur at random on an “infinite length” genotype:
- All mutations are ‘away’. Equivalent to analysing the lineage.
- Reference: B. Derrida and L. Peliti, Bull. Math. Biol. **53**, 355 (1991).
- Question: Is considering a Type - *a “unit” of similar individuals* - useful for understanding diversity?



# Diversity measures

- Measured diversity depends on diversity measure:

- Species Richness:

$$D_{Raw} = \sum_i 1$$

*Sum over species i*

- Simpson Diversity:

$$S_S = 1 - \sum_i p_i^2$$

*Proportion of species i from total population N*

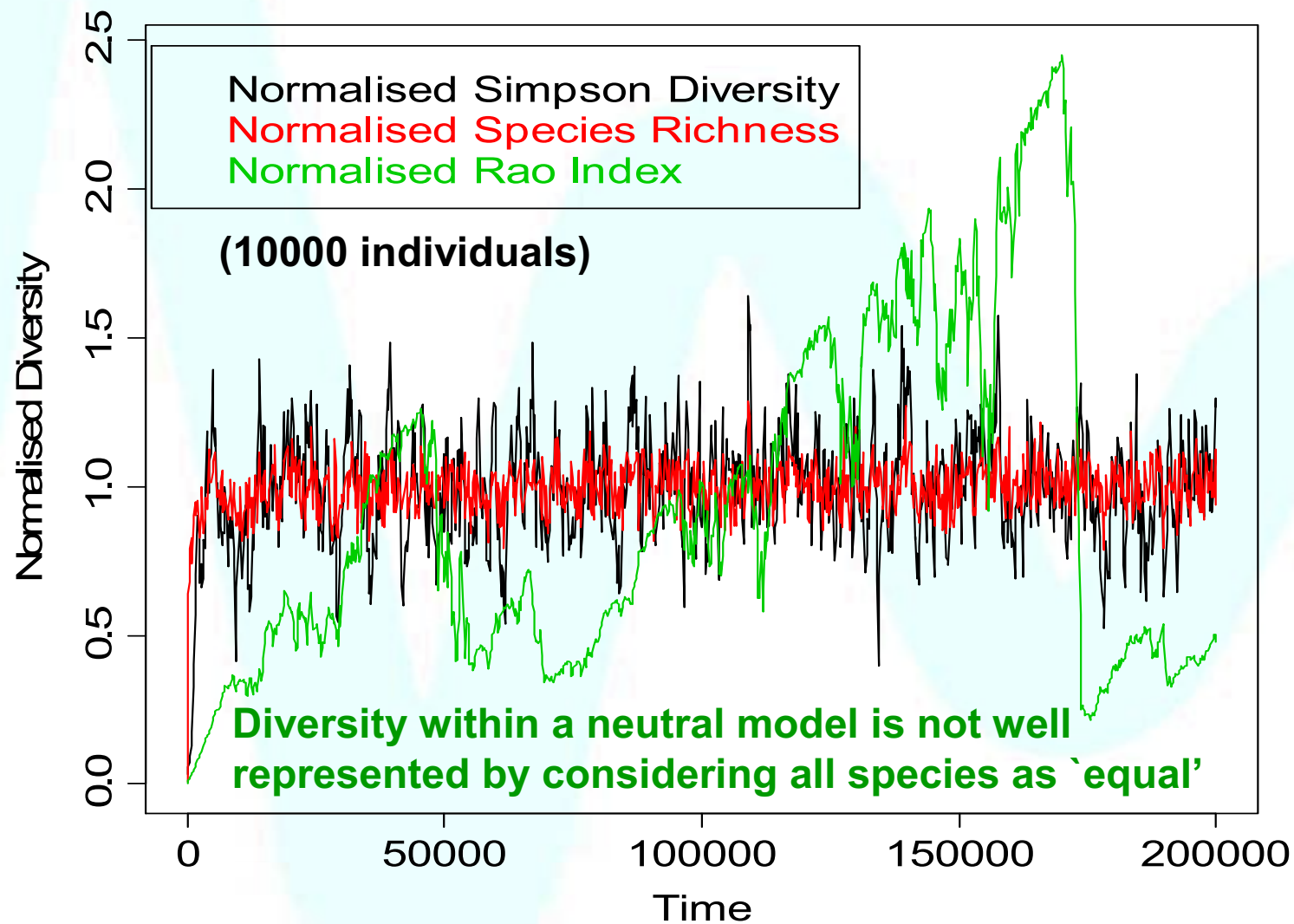
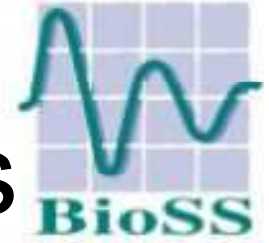
$$D_S = \frac{1}{1 - S_S}$$

- Rao Index:

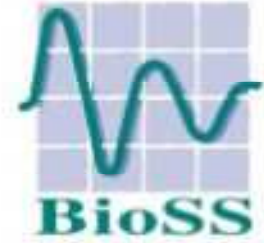
$$D_{Rau} = \sum_{i,j} d_{ij} p_i p_j$$

*"Difference" between types*

# Diversity of asexual lineages

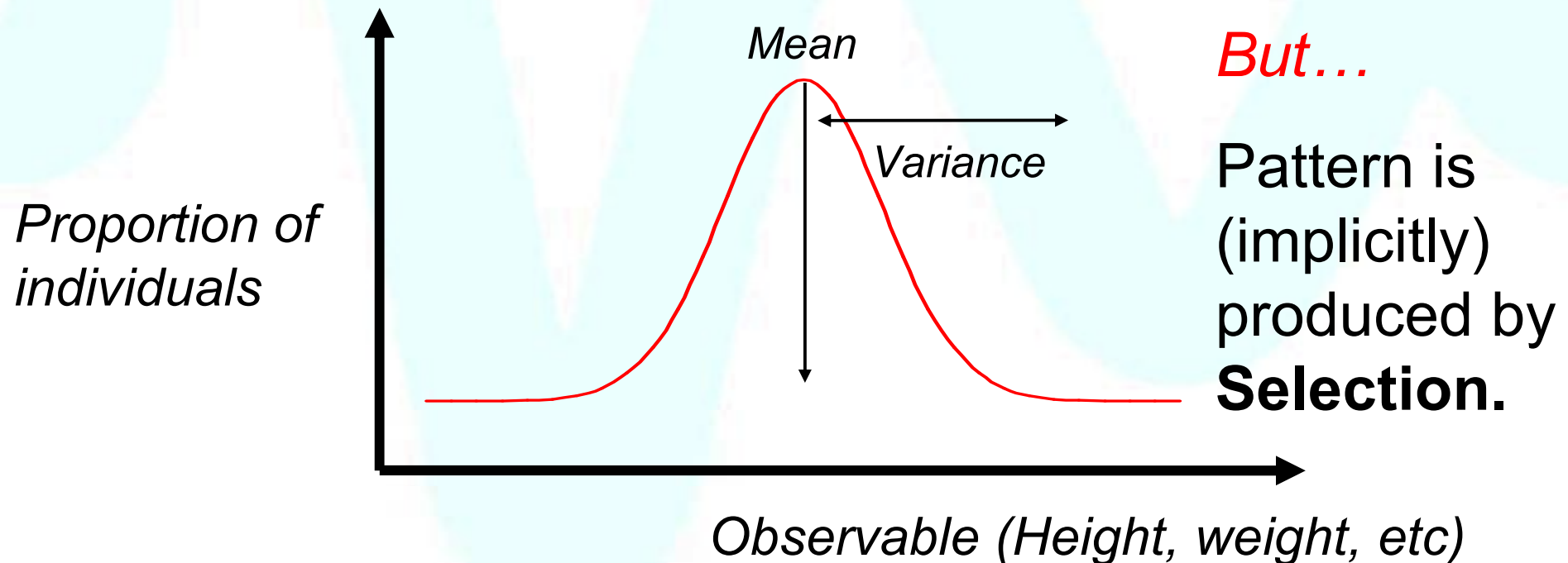






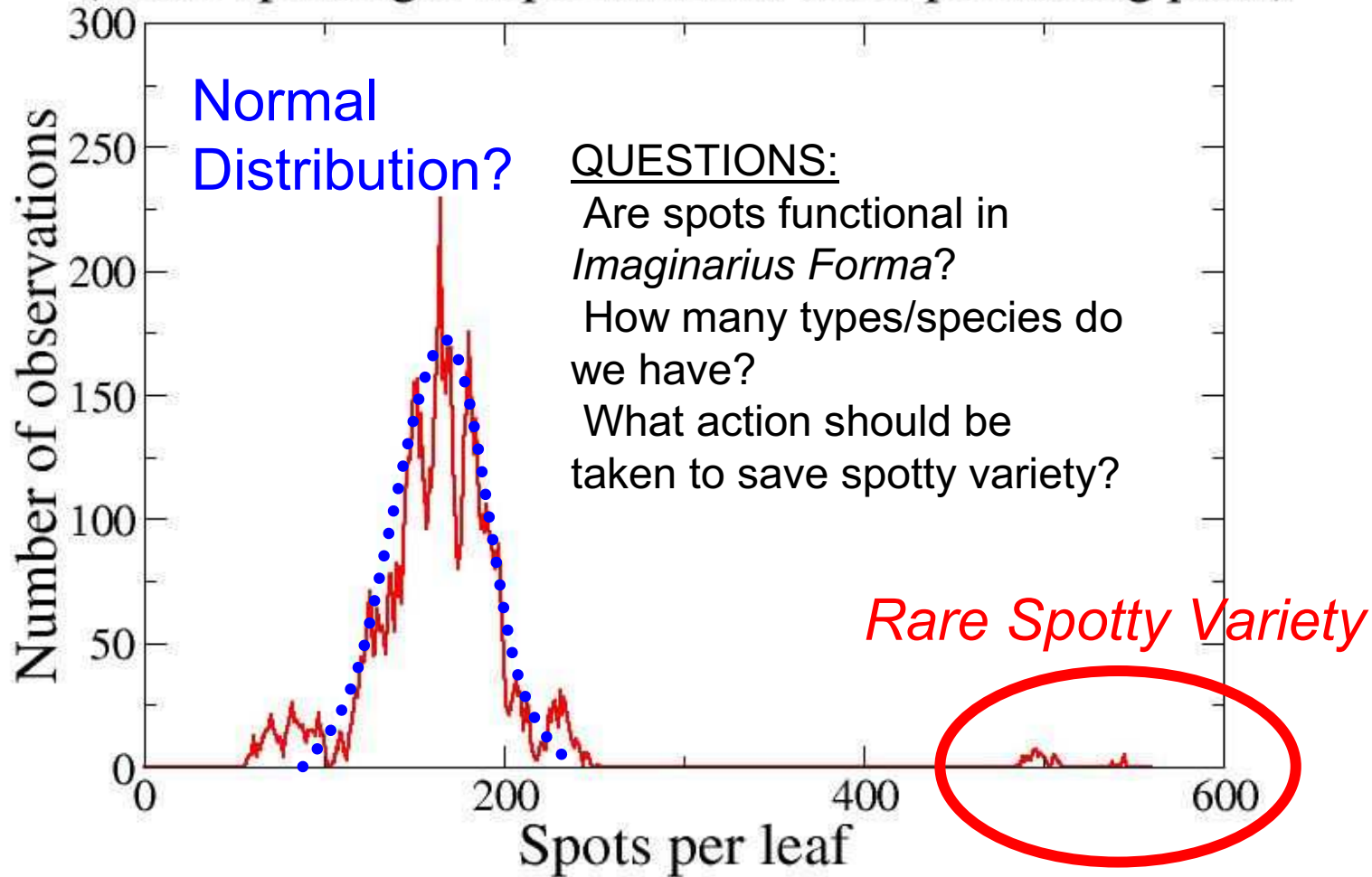
# Phenotype Distribution

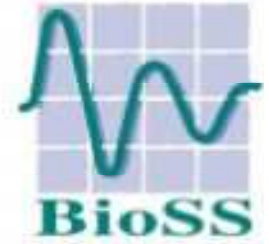
- Consider 1 dimensional case: mutations can be either to the left or to the right.
- Expected pattern is a Normal Distribution:



# Test Problem

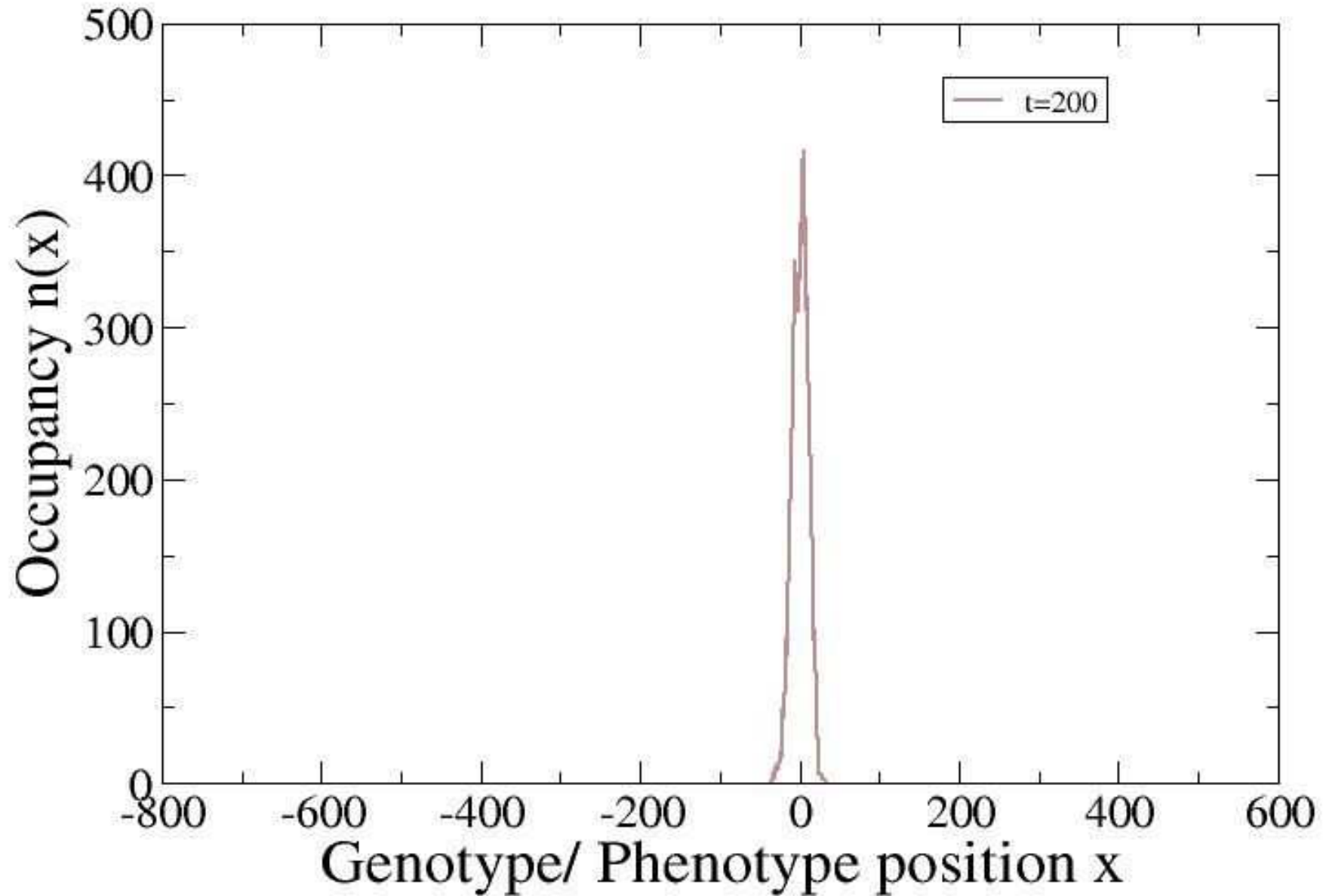
Observations of number of spots per leaf of *Imaginarium Forma*  
(Made up thought experiment for a self pollinating plant)



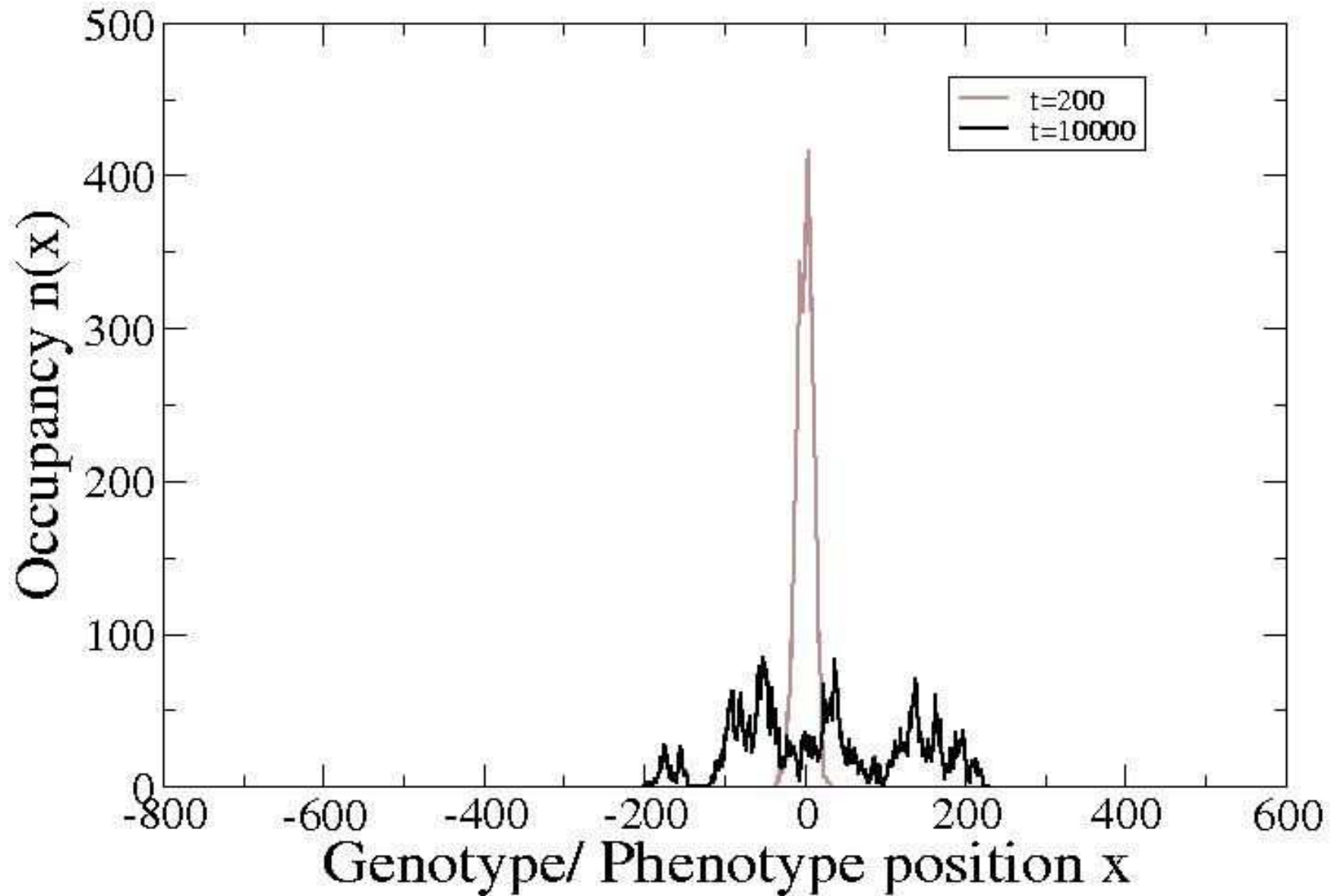


So what does the neutral  
phenotype model do?

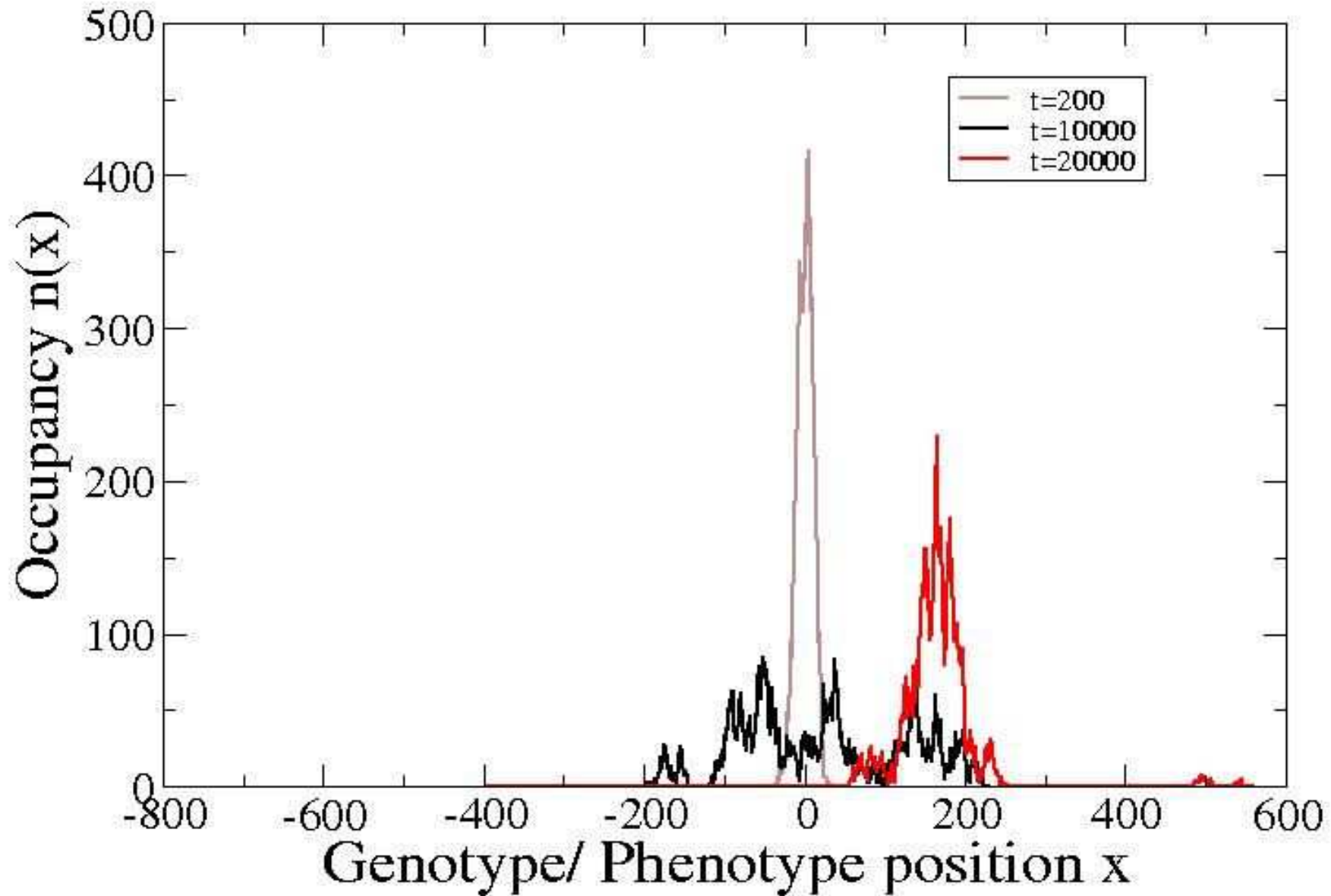
# Evolution of 10000 individuals



# Evolution of 10000 individuals

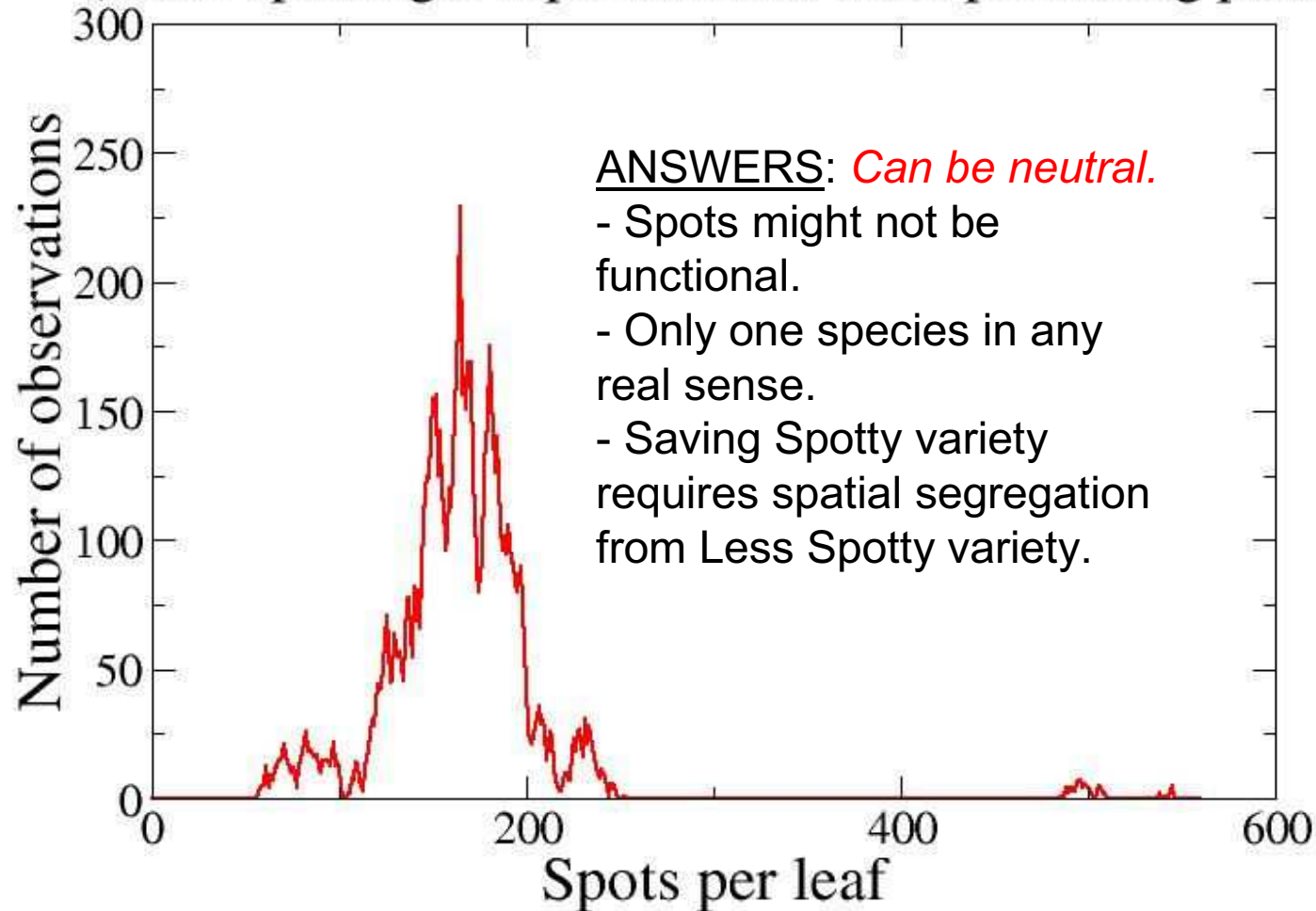


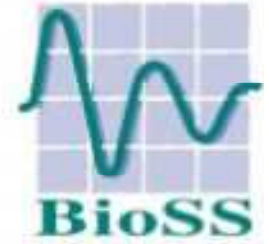
# Evolution of 10000 individuals



# Test Problem

Observations of number of spots per leaf of *Imaginarium Forma*  
(Made up thought experiment for a self pollinating plant)



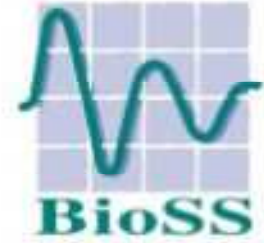


# Solution

- Instead of a “species”, use a conceptual model of a “peak”:
- A Normal Distribution with dynamic mean  $\mu$  and width  $w$ .

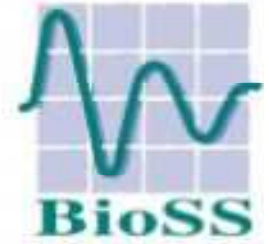
- (1) Select death location  $x$
- (2) Select birth location  $y$ , mutated by 1 with probability  $P_m$
- (3) Remove individual from death location and place at birth location
- (4) Update  $\mu$  and  $w$





# Solution method

- Write down equations for:
  - Changes to the mean,  $\langle d\mu \rangle$ ,  $\langle (d\mu)^2 \rangle$ , ...
  - Changes to the width,  $\langle dw \rangle$ ,  $\langle (dw)^2 \rangle$ , ...
- Take continuous limit to obtain Stochastic Differential Equations
- Solve!



# Solving for the width

$$d(w^2) = \left( p^* - \frac{2w^2}{N} \right) dT + \frac{2w^2}{\sqrt{N}} dW$$

Mutation distance                      Generation time

dW is Random, mean 0

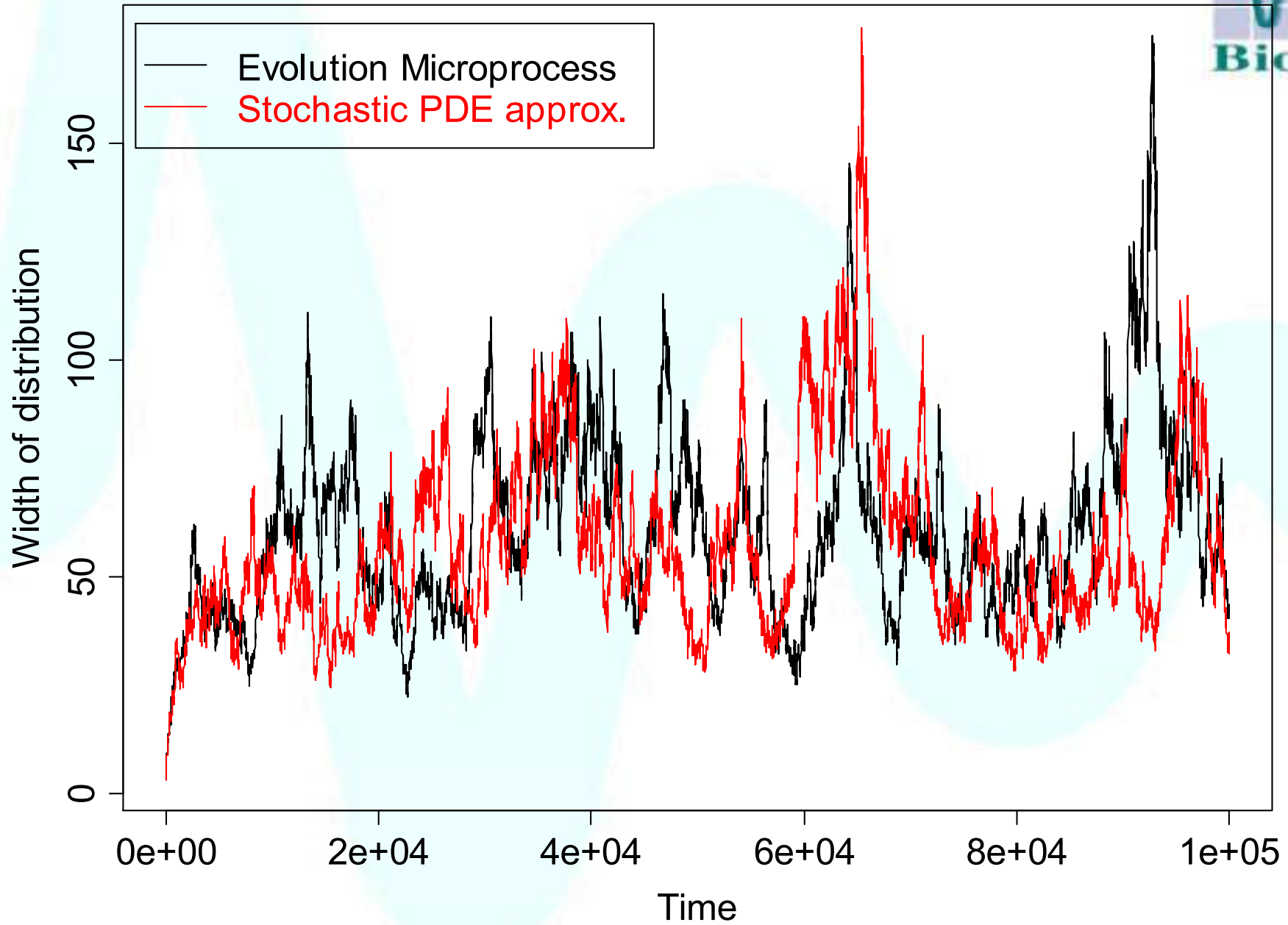
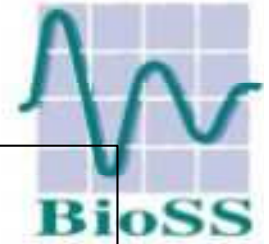
Change in variance (in a timestep) = Deterministic part + Noise part

Solution at steady state:

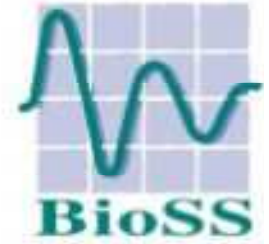
$$p(w)dw = \frac{(Np_m)^2}{2w^5} e^{-\frac{Np_m}{2w^2}} dw$$

Power-law decay at large w

# Evolution in a 1D type space



# Neutral Clustering results



- Mean width:

$$\langle w \rangle = \sqrt{\frac{N p_m \pi}{8}}$$

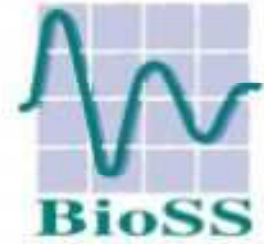
*Fluctuations in  
w also  $\sim N^{0.5}$*

- Drift of Position:

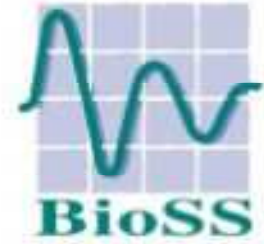
$$\langle x \rangle_{\text{RMS}} = \sqrt{T(p_m + w^2)} \approx \sqrt{\frac{p_m T}{2}}$$

*With time in  
generations...  
 $\langle x \rangle_{\text{RMS}}$  is  
independent of  
N!*

# Neutral evolution results



- Selection produces a stable peak, neutrality produces an *unstable* peak.
- Characteristic peak width, but large fluctuations (*multiple clusters*).
- Evolution speed *independent* of population size!
- *Evolution* is important for *Ecological* questions: *non-trivial* phenotype distributions are probable.
- The **Genotype** Distribution evolves similarly....
- Also relevant for **sexual** populations



# Reference

*Lawson and Jensen:*

***“Neutral Evolution as Diffusion in phenotype space:  
reproduction with mutation but without selection”***

*Physics Review Letters, March 07 (98, 098102)*

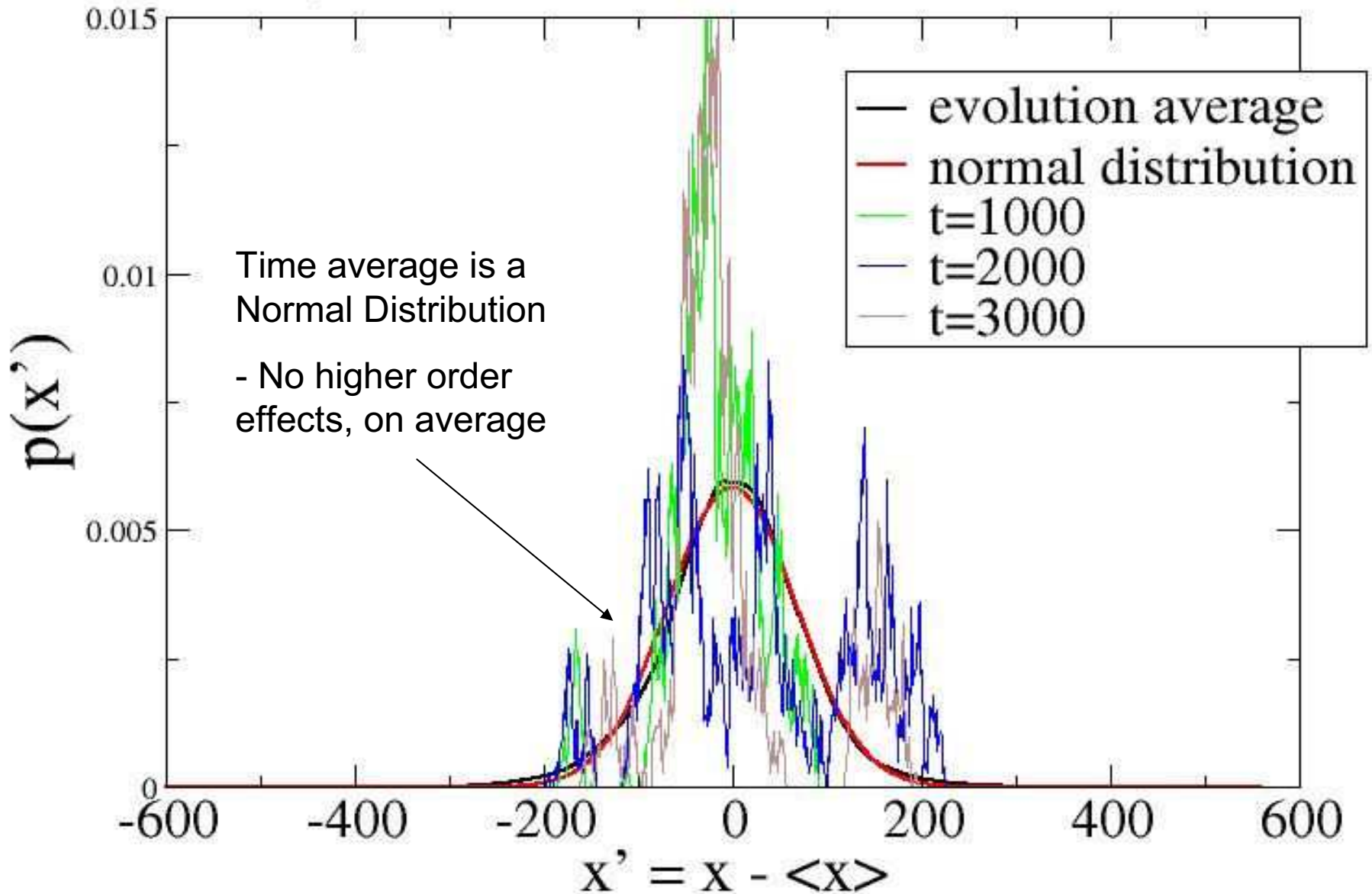
[www.arxiv.org/abs/q-bio/0609009](http://www.arxiv.org/abs/q-bio/0609009)

**Thank you for your attention!**



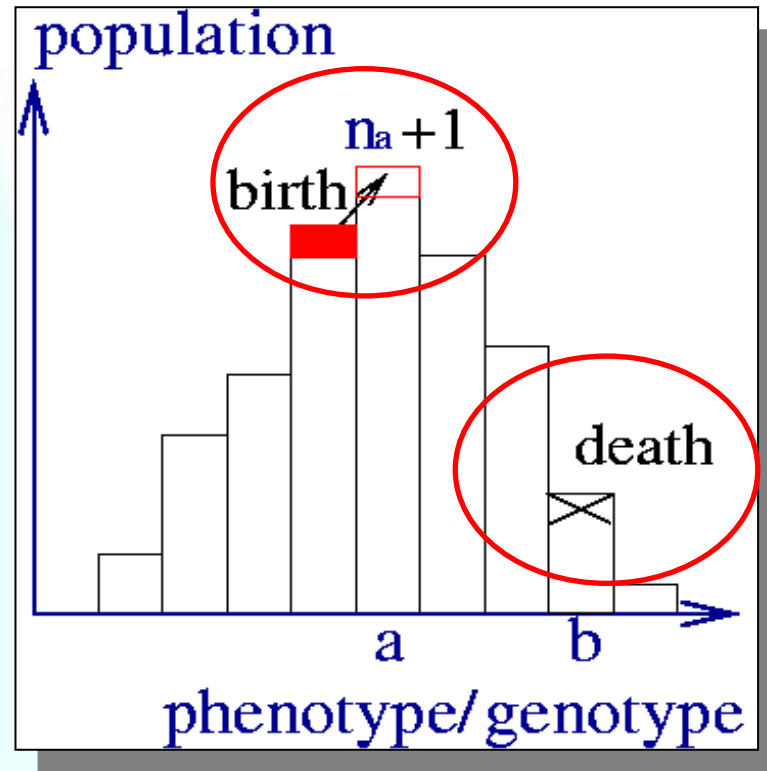
**Cut material**

Time average distribution of evolution process around the mean  
compared with a normal distribution (N=10000)



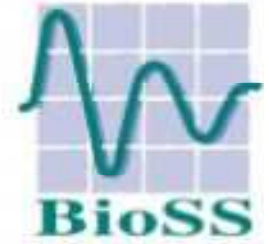


# Solution: first try (1)



$$p_b(a) = \frac{n_a}{N} (1 - p_m) + \frac{p_m}{2} \frac{n_{a+1} + n_{a-1}}{N}$$

$$p_k(a) = \frac{n_a}{N}$$



## Solution: first try (2)

- Consider average behaviour:

$$\langle n_a(t+1) - n_a(t) \rangle = p_b(a) - p_k(a)$$

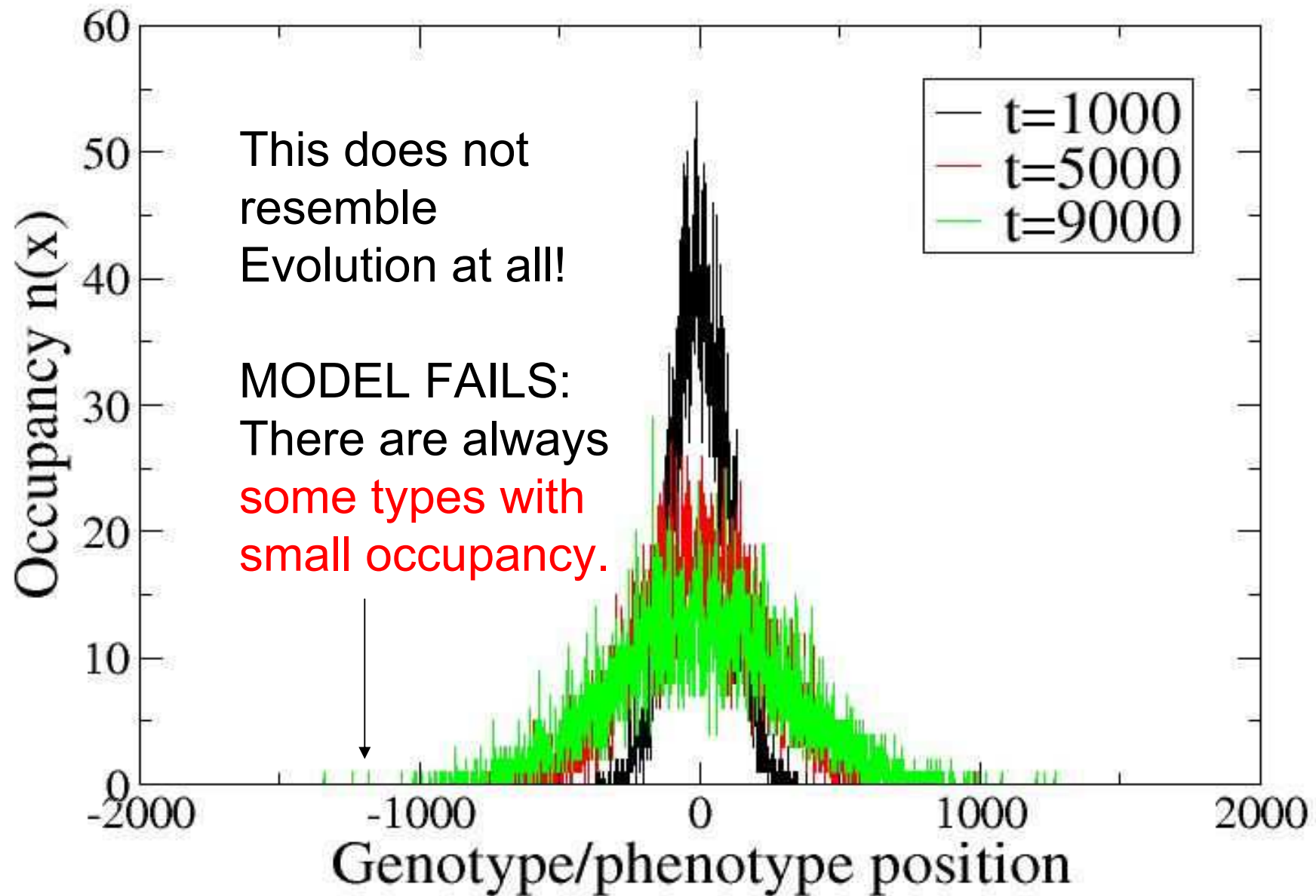
– Expect this to be valid when N large?

- This becomes:

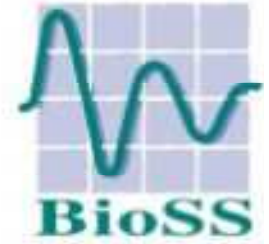
$$\langle n_a(t+1) - n_a(t) \rangle = \frac{p_m}{2N} [n_{a+1}(t) + n_{a-1}(t) - 2n_a(t)]$$

- This is just the diffusion of N particles!

# Diffusion of 10000 particles



# Defining Asexual Diversity



- We want the “number” of different species or types
- Requires definition of species:
  - *Biological Species Concept?*
  - *Phenotypically distinct?*
  - *Genotypic Species Concept?*

## HOWEVER:

- “Species” don't exist – though individuals cluster in type space
- Instead examine the similarity between types