

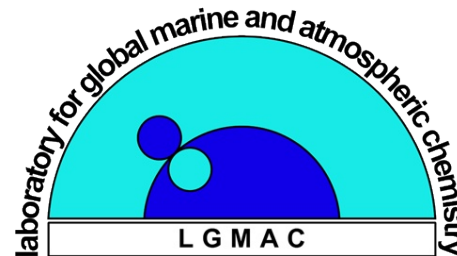
# An evolutionary model of genome streamlining in free-living prokaryotes

Piotr Bentkowski

Hywel T.P. Williams, Thomas Mock, Timothy M. Lenton

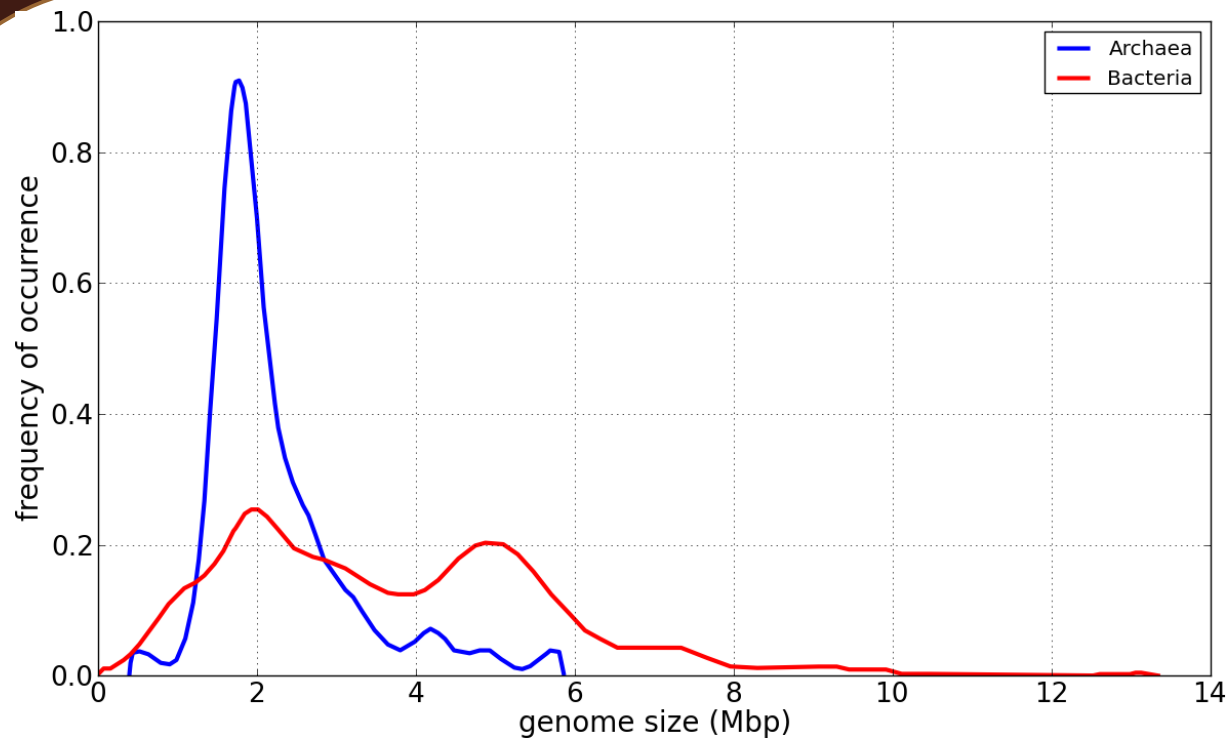
The logo for the University of East Anglia, featuring the letters 'UEA' in a stylized, blue, serif font.

University of East Anglia

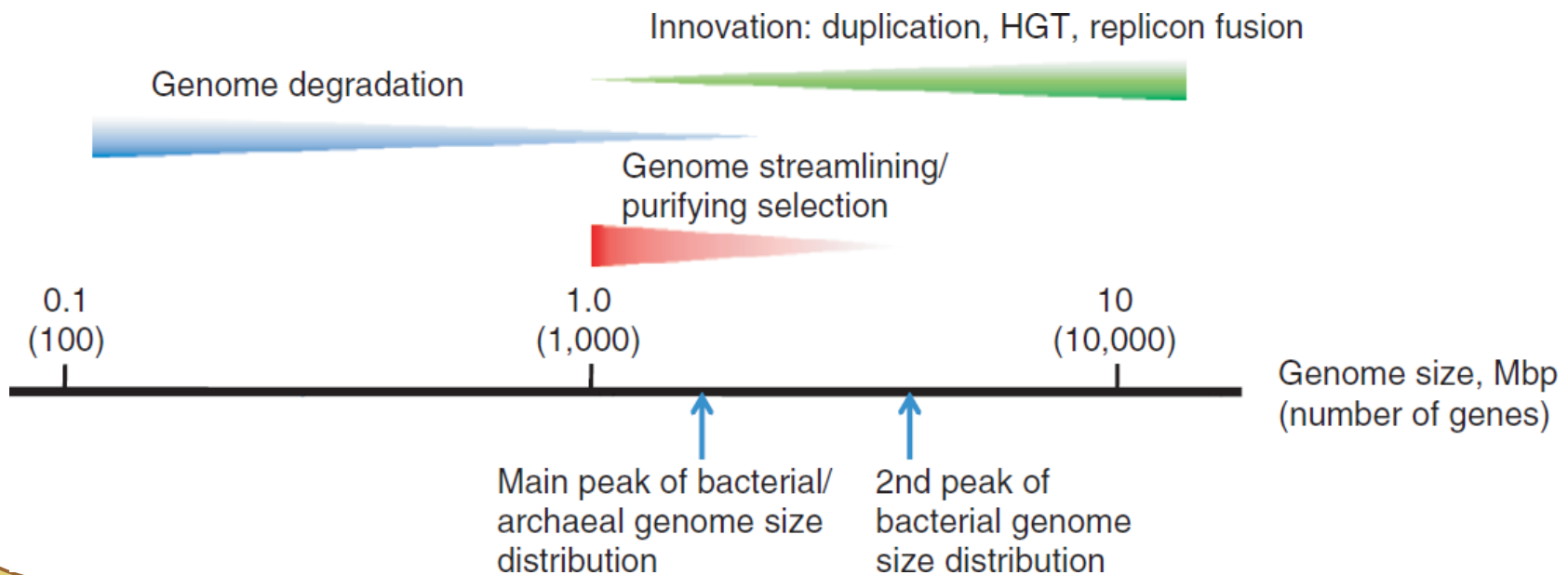


13<sup>th</sup> Congress of the  
European Society for  
Evolutionary Biology

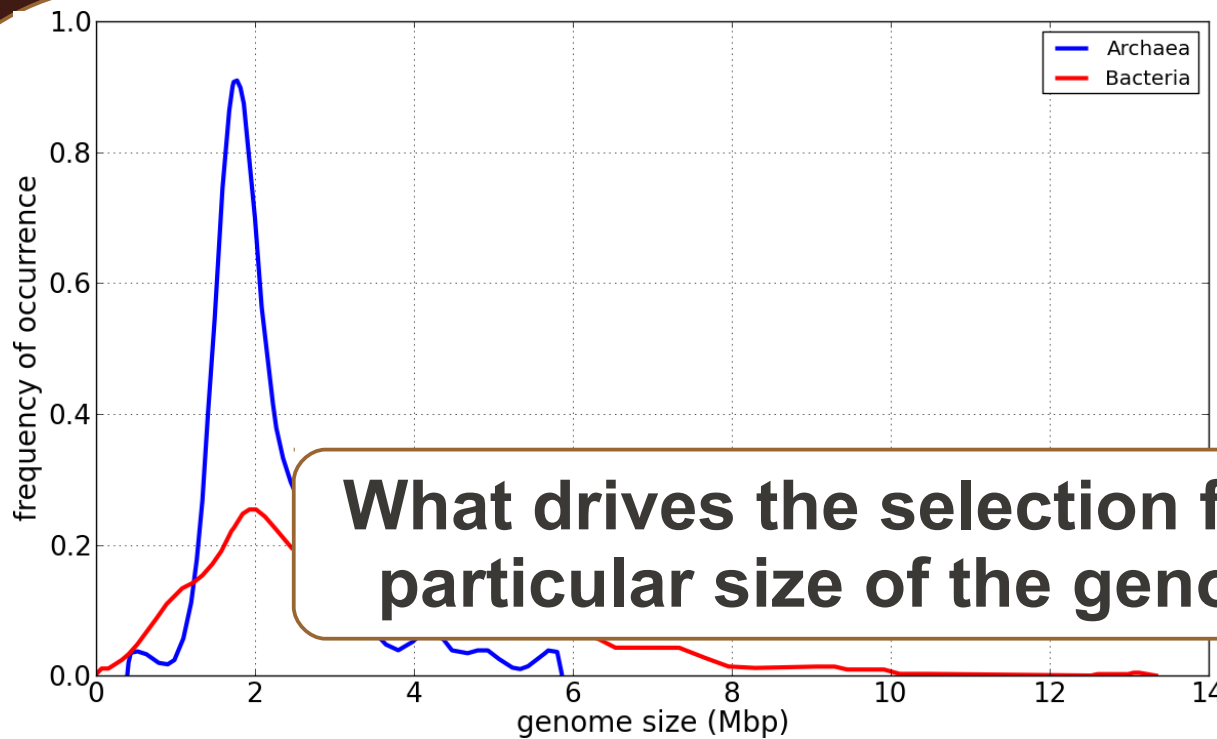
# Background



Distribution of genome sizes in **Bacteria** (red) and **Archaea** (blue) and the principal forces of evolution in prokaryotes and their effects on archaeal and bacterial genomes. (both modified from Koonin and Wolf, 2008)



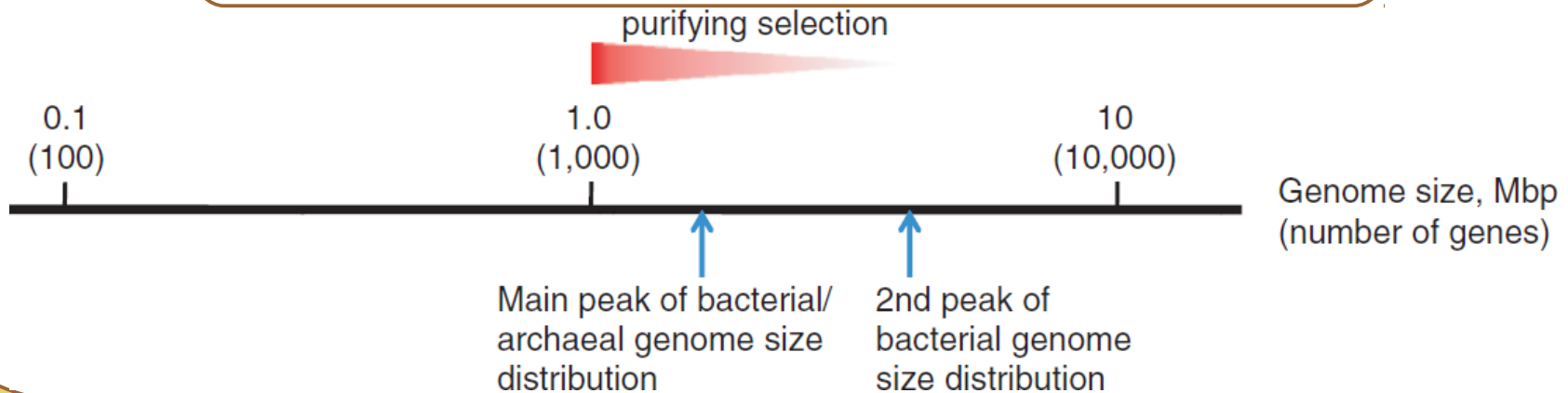
# Background



Distribution of genome sizes in **Bacteria** (red) and **Archaea** (blue) and the principal forces of evolution in prokaryotes and their effects on archaeal and bacterial genomes. (both from Koonin and 08)

**What drives the selection for the particular size of the genome?**

**Does the genome size impact the adaptive potential of a prokaryote?**



# Hypothesis

---

Stable abiotic conditions enhance selection pressure to reduce the cost of living and the cost of reproduction – cells will have smaller and cheaper genomes.

Unstable abiotic conditions bring more challenges which can happen and the cell has to be prepared to tackle them – cells will have more genes and will spend more resource on using them.

To tackle the environmental changes, evolution rather rearranges existing solutions than invents brand new ones. Populations with strongly streamlined genomes will be more valuable to extinctions.

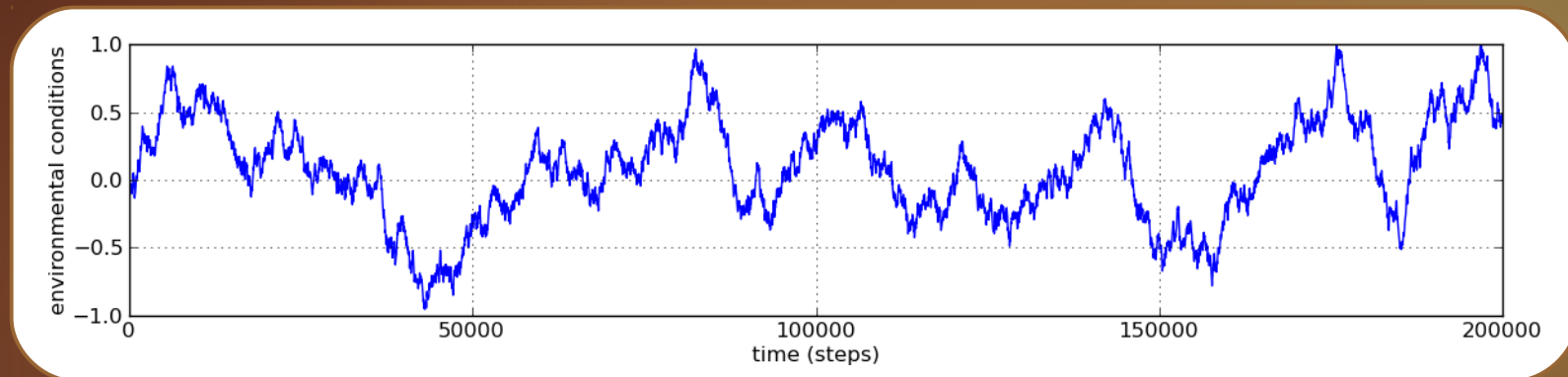


# Methods – basics of the model

---

Cells compete over a single resource represented as a big-but-finite real number.

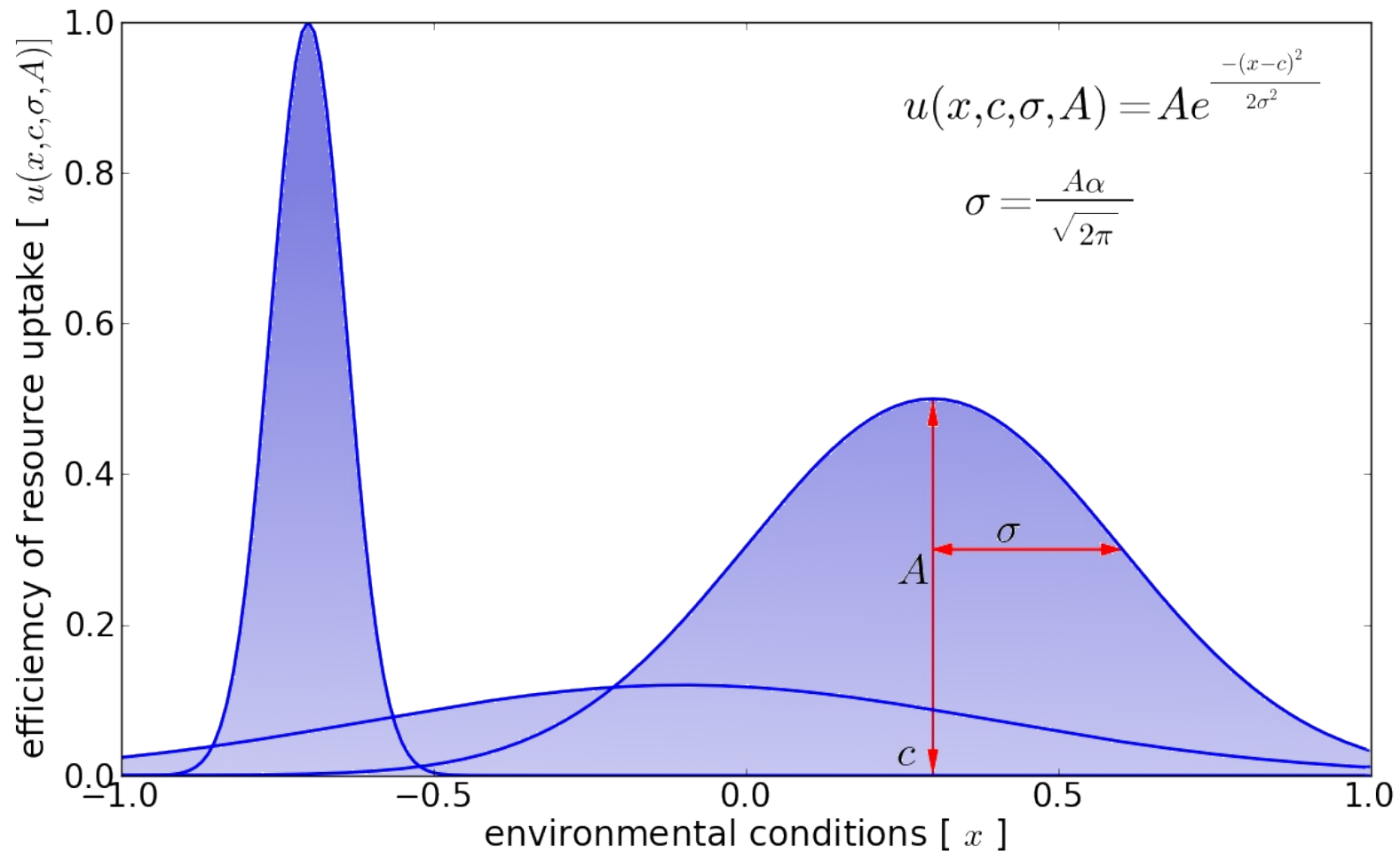
Abiotic condition is one dimension space in the range between -1.0 and 1.0 (think of it like e.g. pH). State of the environment changes in time by performing a bounded 1D random walk with a defined maximal length of a step (called *turbulence* in this research).



The performance of a cell depends on the quality and the number of metabolic genes it has. Each gene generates a cost (replication, transcription, translation,...)

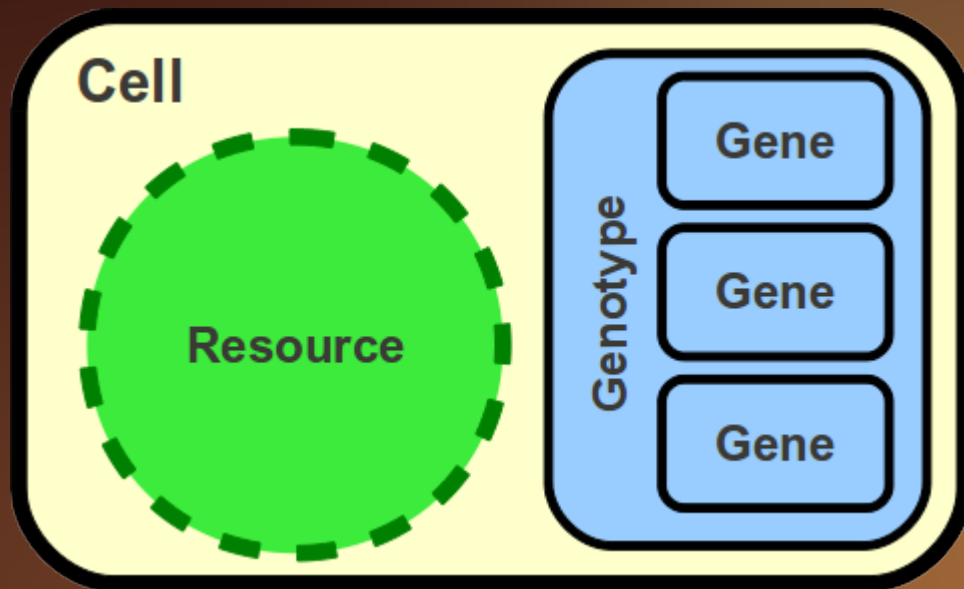
Genome having  $N$  'metabolic' genes will also have an overload of  $N^2$  regulatory genes which generate costs too (genome scaling laws: e.g. van Nimwegen, 2003).

# Methods – genes and genotype



An example of a genotype having three genes. Note, that the surface under the curve is fixed. This is a trade-off between efficiency and generality which prevents 'supergenes' from occurring. Best uptake a cell can get for the given environmental condition is the value of the best gene in that location. Not a sum, nor a product of genes.

# Methods – cell



## Mutations

- deletion – gene is removed
- duplication – gene is copied and we have two same genes
- modification – gene's Gaussian curve params are changed

## Cell – the agent of the model

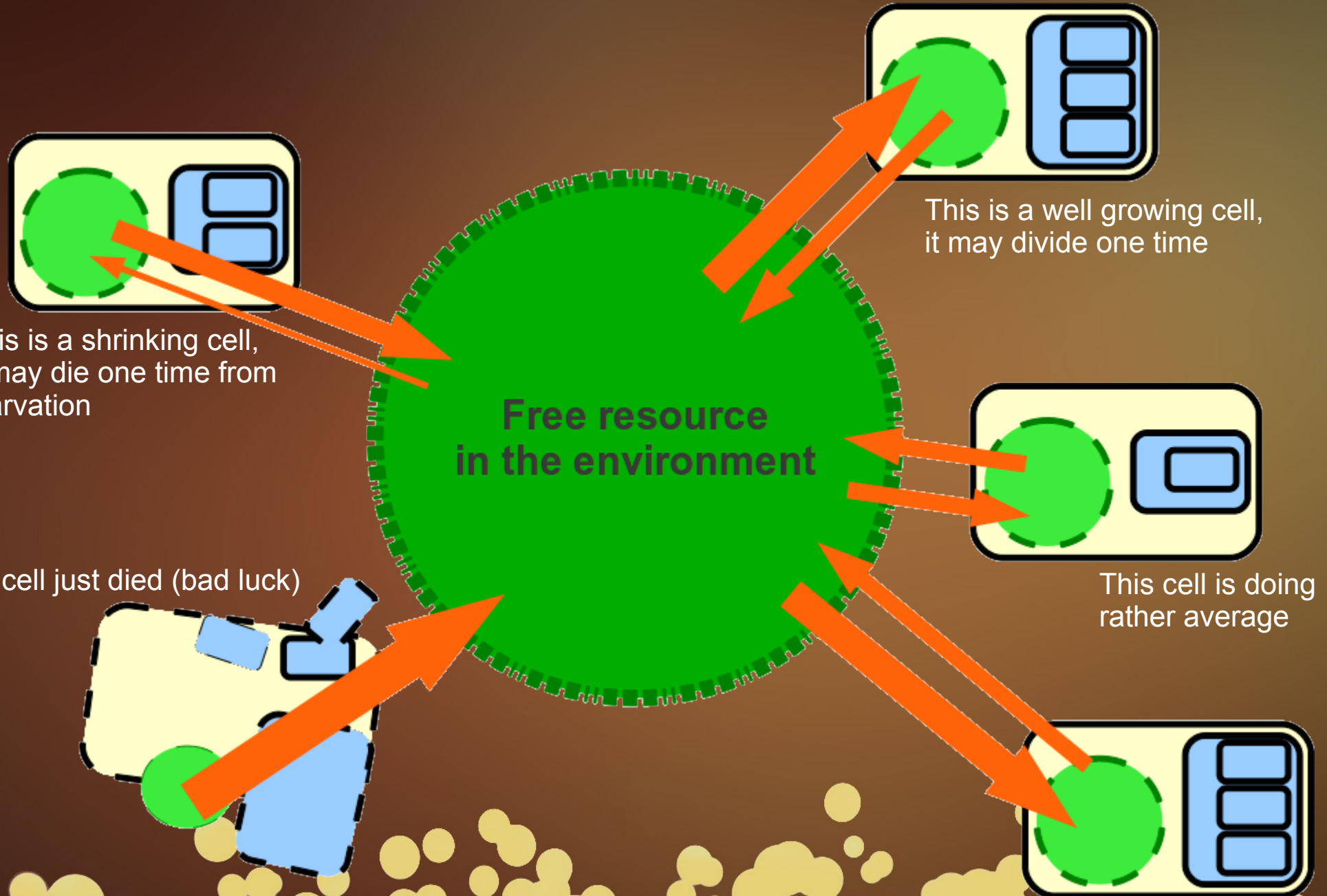
**It has:** its internal pool of resource, genotype with genes (*as seen before*)

**It does:** uptake resource, give resource back to the environment (*gene expression cost, breathing*), divide (*if the internal pool is big enough*), die (*if the internal pool is too small*)

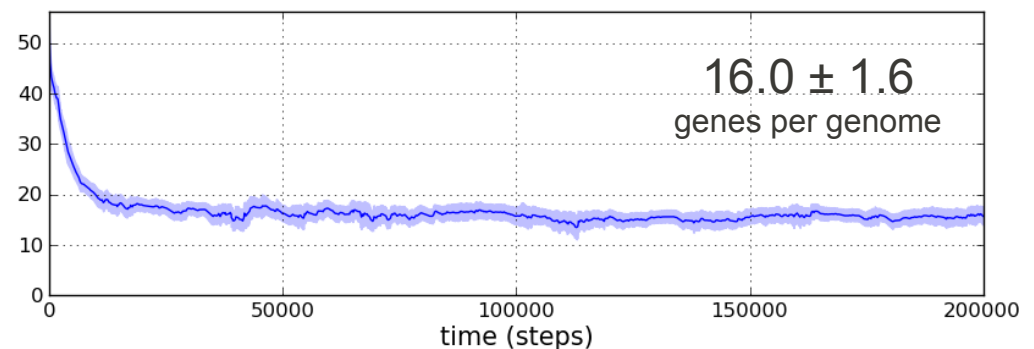
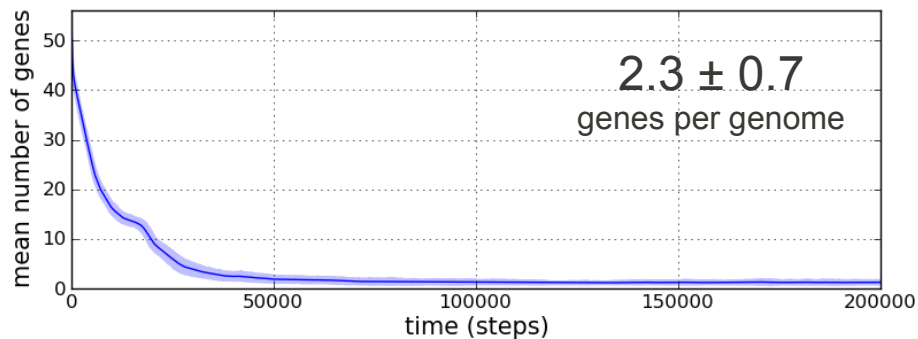
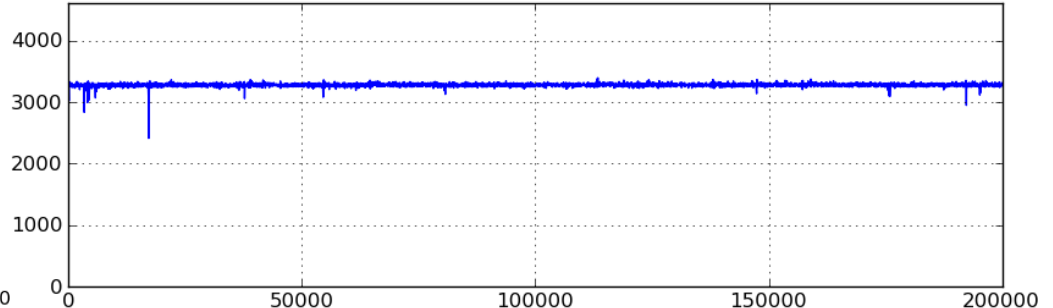
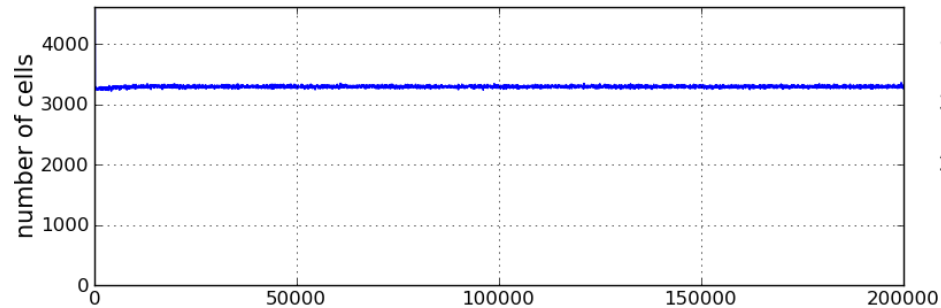
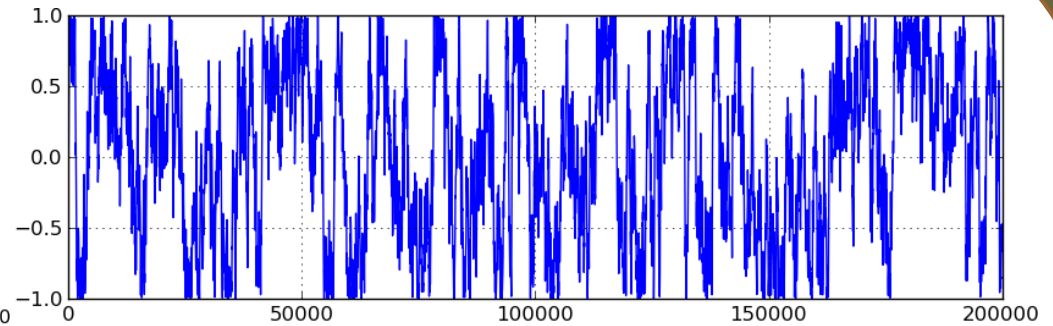
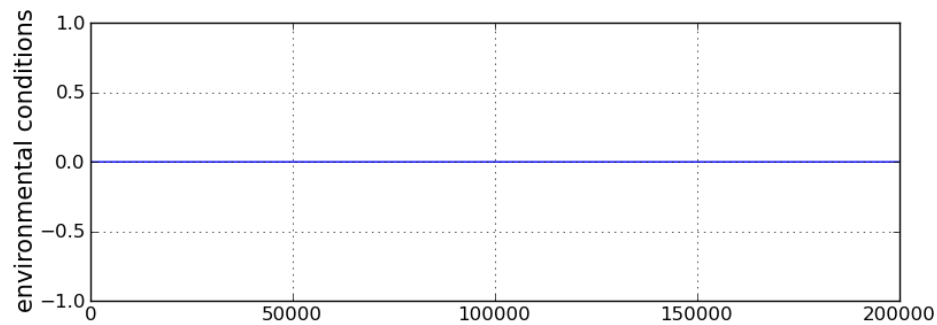
**It is subject to:** probability of mutation at division, probability of random death at each time step

If cell dies, all its resource are returned to the environment

# Methods – ecosystem

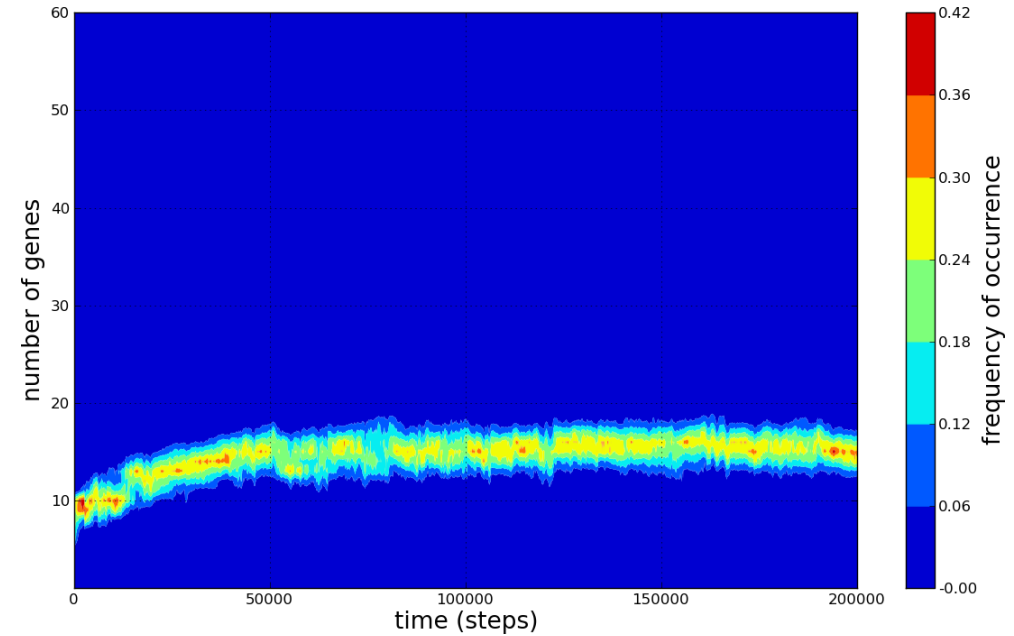
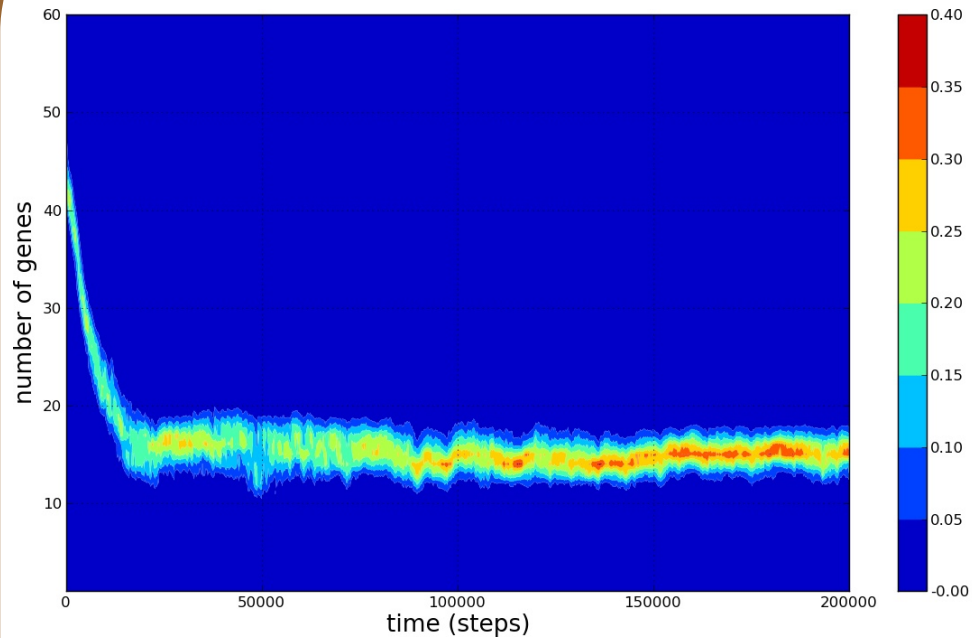


# Results – different turbulence



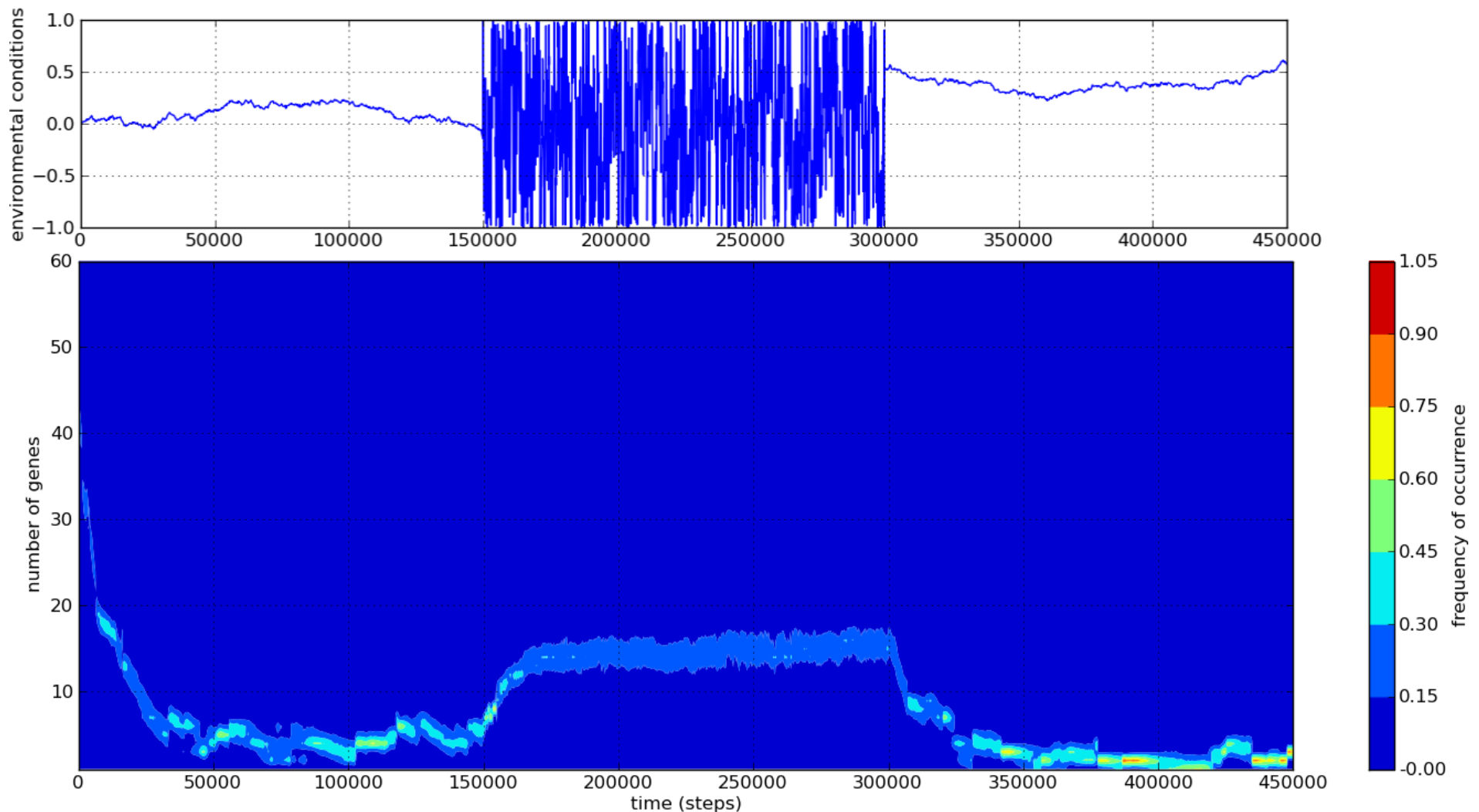
Two simulations initialized with high number of genes. System with no turbulence is on the left, the one with turbulence of 0.1 (maximal length of a step in the random walk is 0.1 per time step) is on the right. Both system reduce gene number and stabilize. Light blue shade in the lower panels is STD of the number of genes in each time step.

# Results – different genome size at initialisation



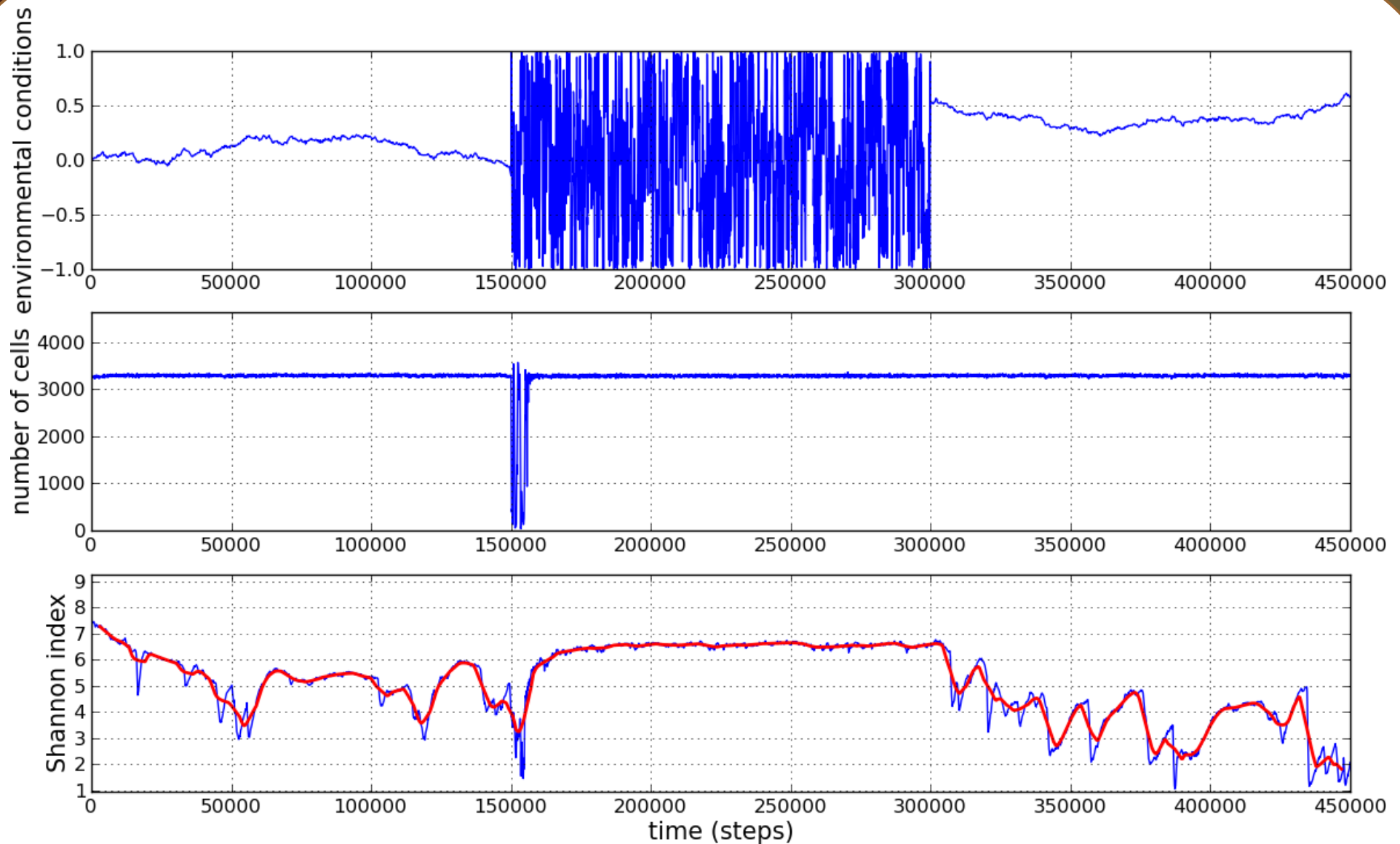
Two simulations initialized with populations of cells with high (left) and low (right) number of genes in the genomes. Both simulations reach a same stable number of genes: for high number of genes at the beginning it finally is  $14.8 \pm 1.5$  genes per genome (mean  $\pm$  STD), for low it is  $14.7 \pm 1.7$  (mean  $\pm$  STD). Both values are averaged over time steps from 100,000 to 200,000 for respective simulations.

# Results – system can adapt to shift in turbulence regime



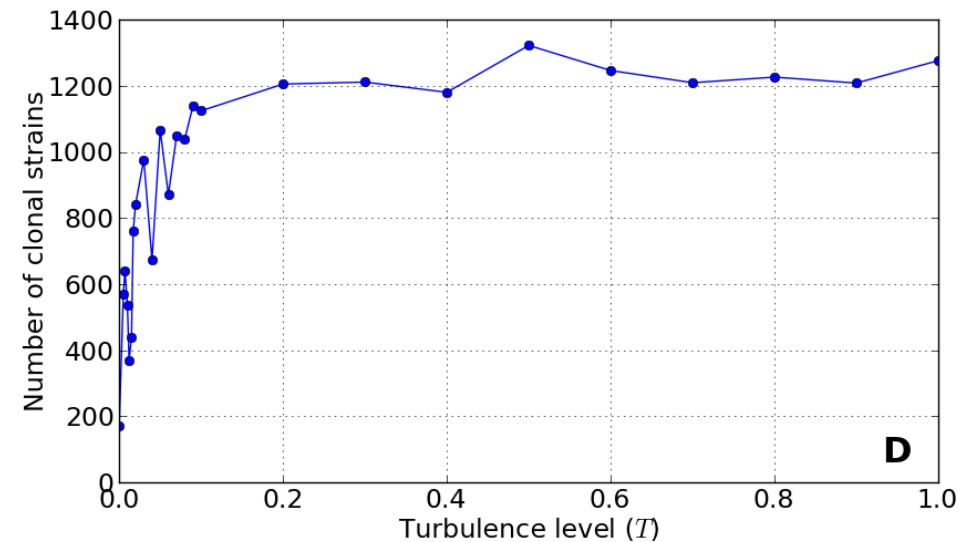
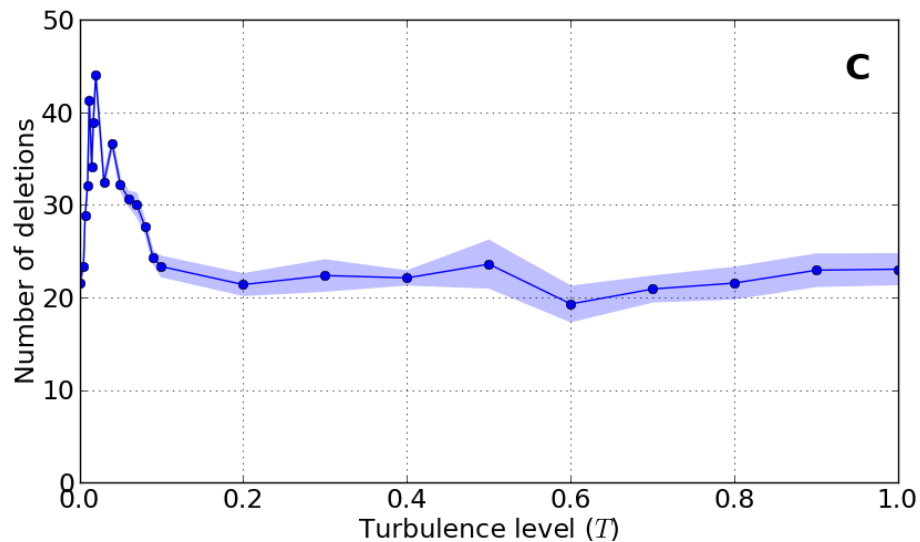
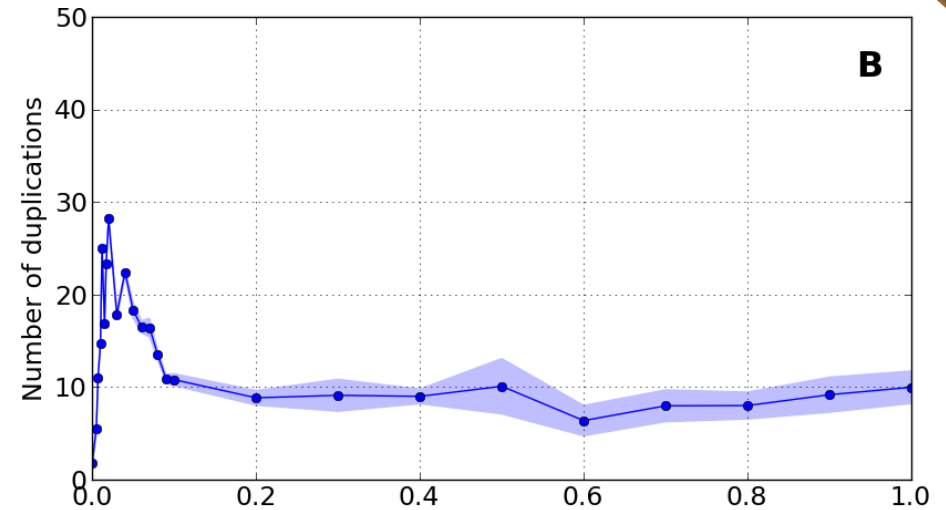
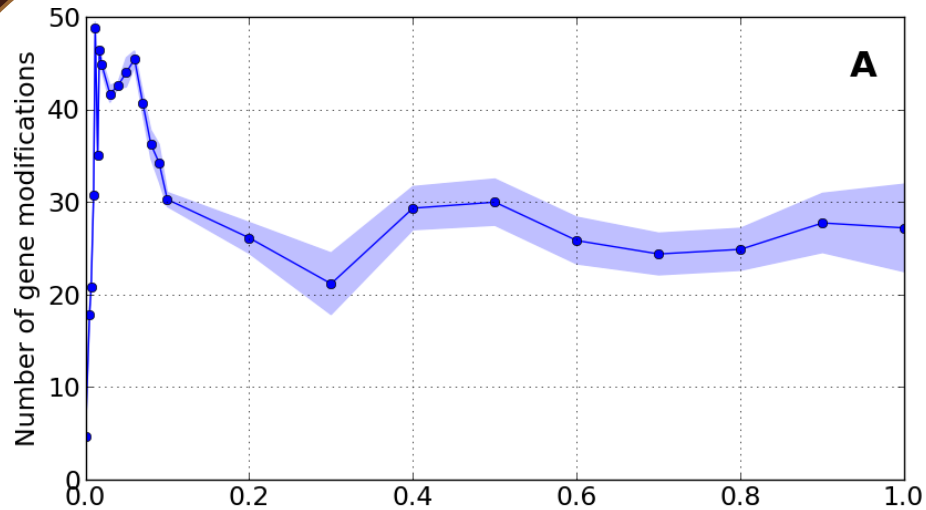
System has the ability to respond to shift in the turbulence regime (three modes on the upper panel) by changing the gene number in the population. However...

# Results – system can adapt to shift in turbulence regime



... the shift from a low turbulence (max length of a step in random walk is 0.002) to a high turbulence (max step is 0.2) triggers a series of a deep extinction events.  
Shannon index (diversity of clonal strains) – lower panel blue line, stepping mean – red line.

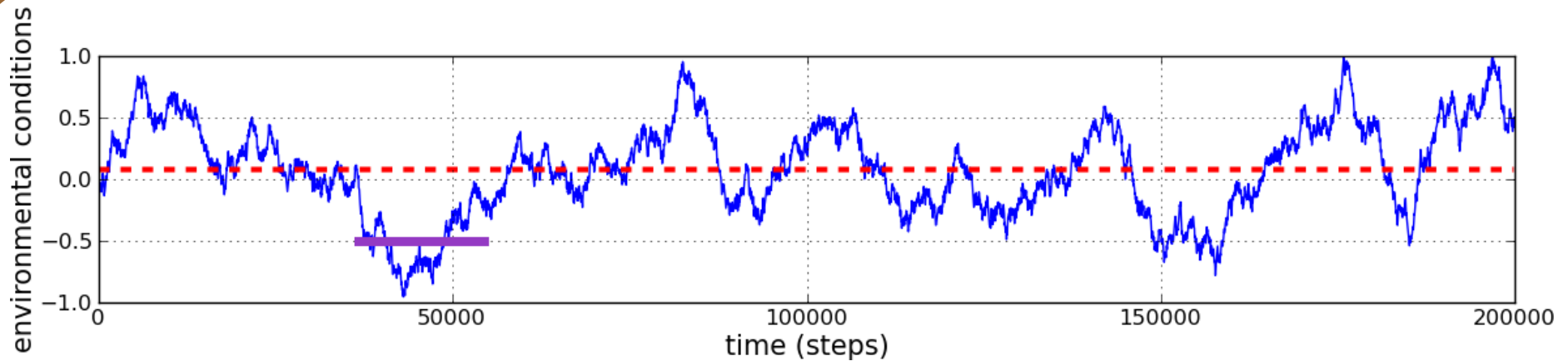
# Results – number of mutations



Number of mutations per 100,000 time steps per clonal strain (panels A, B and C) are the highest for systems with a moderate turbulence. Each dot represent the mean number of mutations per one cell of one clonal strain; shaded area is the STD.

This rise is not accompanied by the rise in the number of clones in the populations (panel D).

# Results – number of mutations

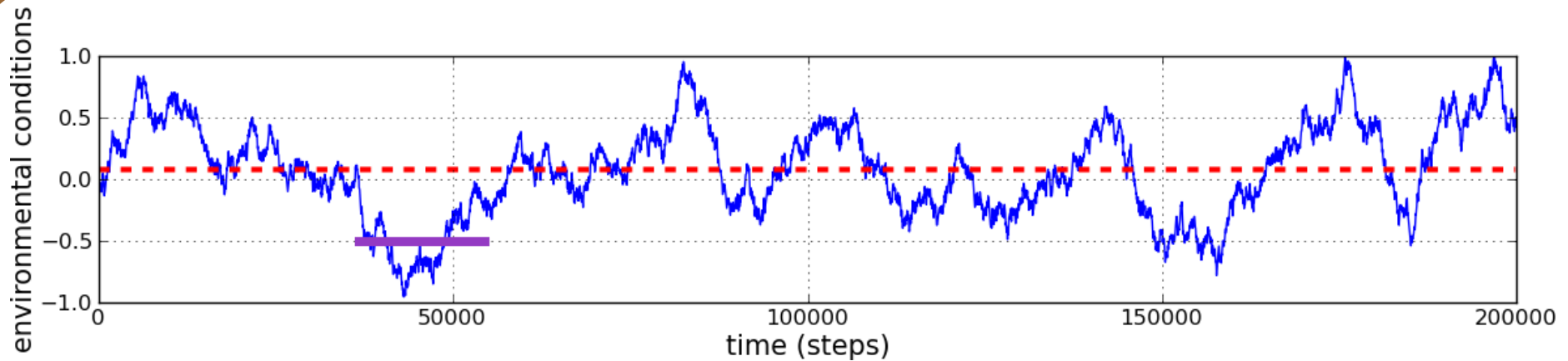


**Red dashed line** is the grand mean of the environmental conditions; **purple bar** is the mean of the environmental conditions within the size of the frame equal to cell's life expectancy.

Let us consider the average absolute difference between the grand mean of the environmental conditions and the mean of the environmental conditions over cell's life expectancy:

$$\bar{D}_f = |\bar{x}_f - \bar{x}|$$

# Results – number of mutations

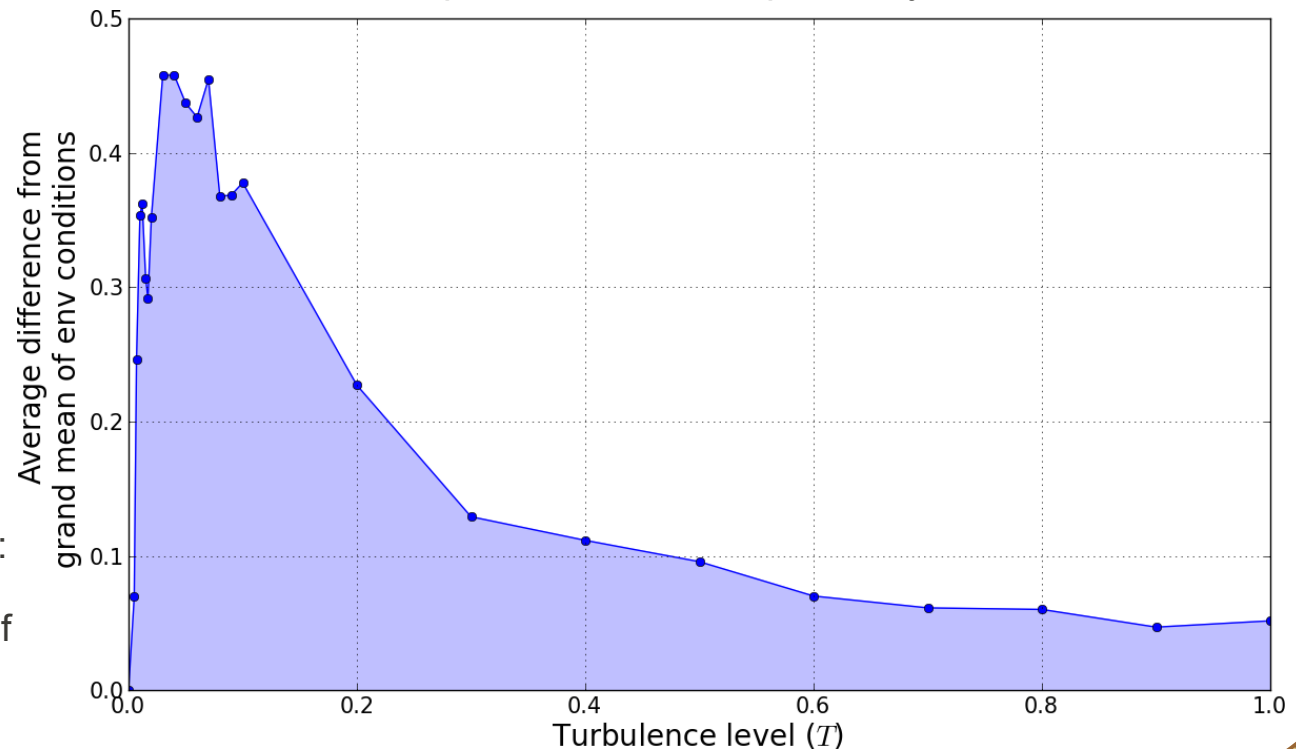


**Red dashed line** is the grand mean of the environmental conditions; **purple bar** is the mean of the environmental conditions within the size of the frame equal to cell's life expectancy.

Let us consider the average absolute difference between the grand mean of the environmental conditions and the mean of the environmental conditions over cell's life expectancy:

$$\bar{D}_f = |\bar{x}_f - \bar{x}|$$

There are two important time scales:  
**Short** – in which population adapts  
**Long** – which regulates the speed of the evolutionary process



# Summary

---

The distribution of genomes sizes in prokaryotes can be explained in the light of a trade-off between the need of having a small genome to live cheap and the need to have all the necessary tools to mitigate changing environment. We think our model is a good reason to lunch experimental research in this direction.

There are two ways environment can impact the evolution process:

***Turbulence*** – rate of change within organism's life span to which population adapts (short time scale)

***Predictability*** – what is the probability that the current biology of a species will meet the future demands of the environment (long time scale)

Population with strongly streamlined genomes have less capacity to tackle radical environmental change then population with larger genomes.



# Thanks for your attention!

---

This research was conducted at the School of Environmental Sciences, University of East Anglia (Norwich, UK) as a part of a PhD project. There is more to show...

Many thanks to the *Earth System Modelling Group* members: Tim Lenton, Hywel Williams, Stuart Daines, Jim Clark, Valerie Livina, Martin Johnson, Sudipta Goswami, Richard Boyle, Phil Underwood, Ben Mills... and to Thomas Mock and Cock van Oosterhout.

