

A model of genome streamlining in free-living prokaryotes

Trade-off between need of mitigation of instability of environmental conditions and the genome's expression and maintenance costs

Piotr Bentkowski¹, Hywel T. P. Williams², Thomas Mock¹, Timothy M. Lenton¹

¹ School of Environmental Sciences, University of East Anglia, NR4 7TJ, Norwich, UK

² School of Computing Sciences, University of East Anglia, NR4 7TJ, Norwich, UK

P.Bentkowski@uea.ac.uk



Genome streamlining

Q: What are the rules regulating genome size for a given species?

Q: What are the consequences of genome streamlining for species evolution?

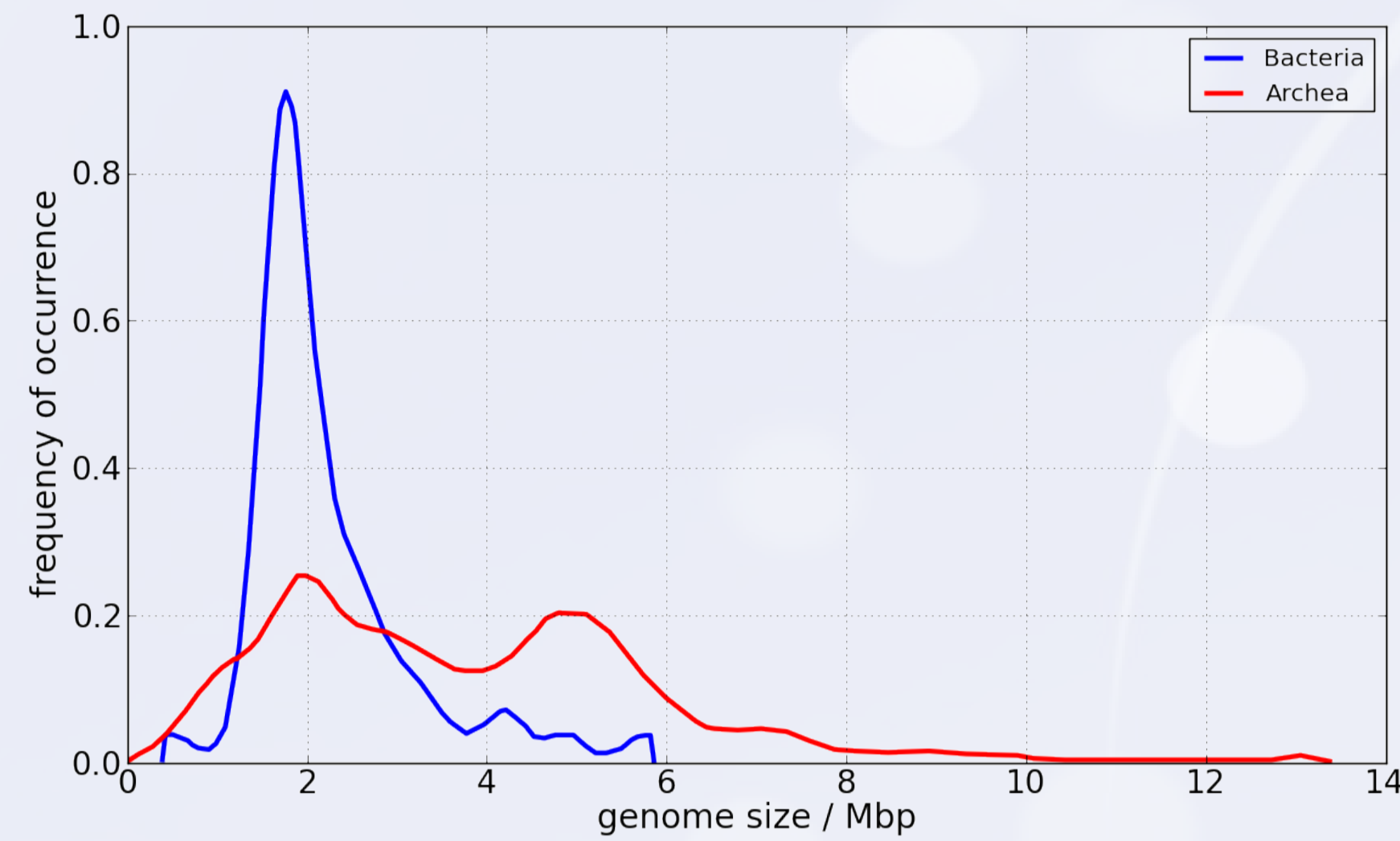


Fig.1. Distribution of genome sizes among Bacteria and Archea (Koonin & Wolf, 2008)

Hypothesis

In stable conditions less genes are required to respond to the environment. Genes which do not bring adaptive benefits cause unnecessary costs and are lost. When the environment is turbulent, organisms need broader response range, thus they require more genes which include all the necessary metabolic pathways. More turbulent environments will support genomes which are bigger and thus 'more expensive', but they insure organisms against the majority of possible changes.

Assumptions

- I. A single gene has a direct impact on the organism's fitness.
- II. Selection pressure recognises only the phenotypic outcome of genes.
- III. The cost of maintaining DNA calculated per gene is fixed.
- IV. There are genes which cannot be removed from a free-living organism's cell.
- V. The number of regulatory genes scales to number of 'metabolic' genes with n^2
- VI. The amount of non-coding DNA is negligible in prokaryotes.

Model's general properties

An agent-based modelling approach was used. Population consists of cells (agents), each having its own unique genotype and own resources which are exchanged with the environment.

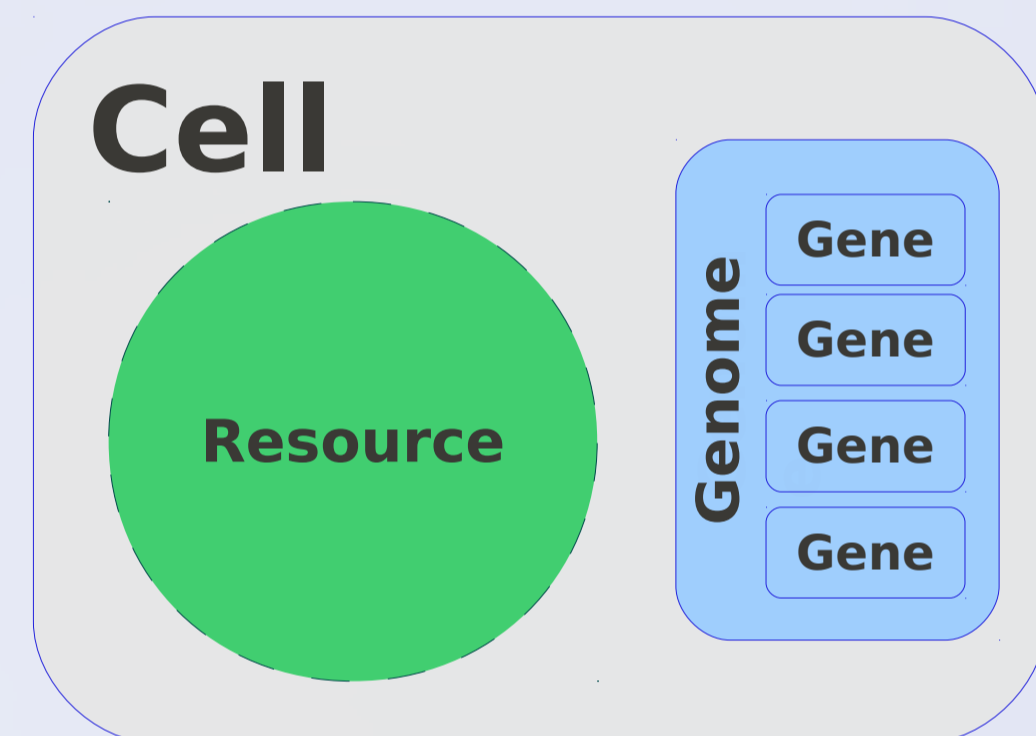


Fig.2. Scheme presenting the construction of one single agent used in our approach. Objects (Solid line) are embedded in each other. Individual resource of the cell (dashed border) changes during cell's life determining about its death or reproduction.

Genome size optimization

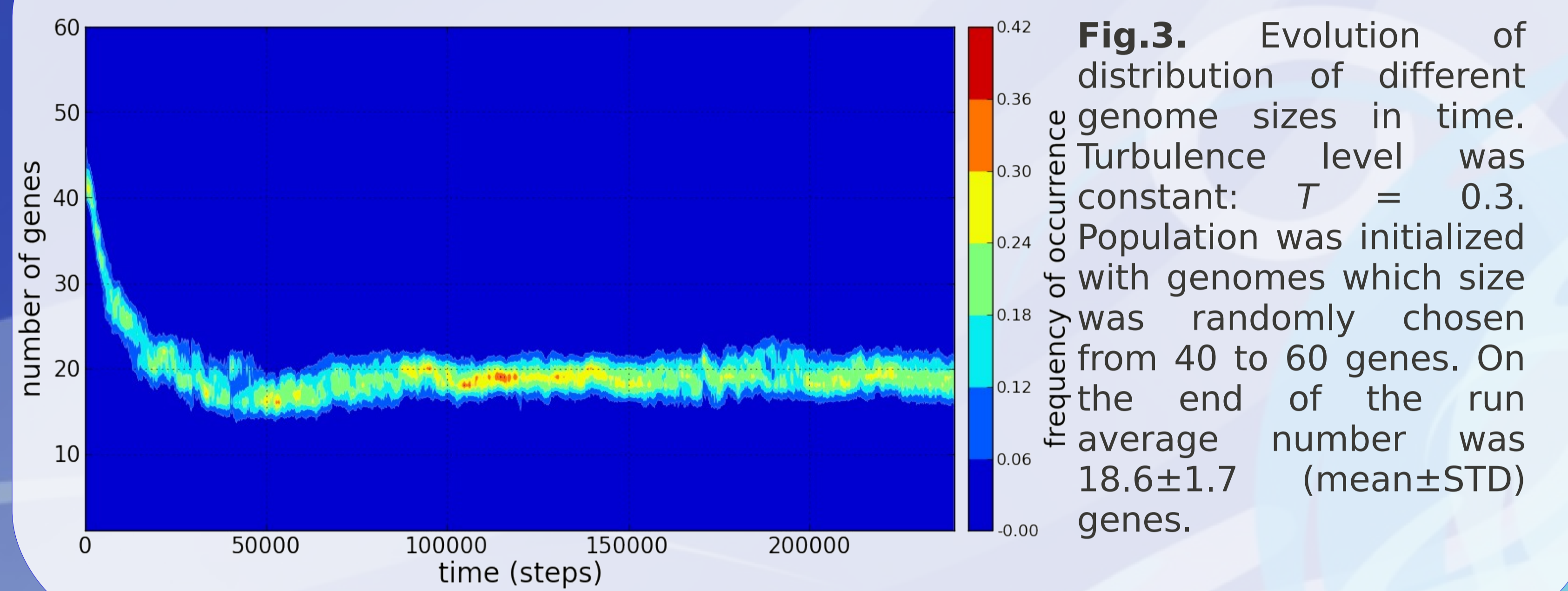


Fig.3. Evolution of distribution of different genome sizes in time. Turbulence level was constant: $T = 0.3$. Population was initialized with genomes which size was randomly chosen from 40 to 60 genes. On the end of the run average number was 18.6 ± 1.7 (mean \pm STD) genes.

Genotype representation

Genes are represented by Gaussian curves (described by 3 parameters) defined in a 1D Hutchinson niche space. Genotype is a set of n genes. A , c and n are subject of evolution. σ is defined by fixed α to prevent 'supergenes'.

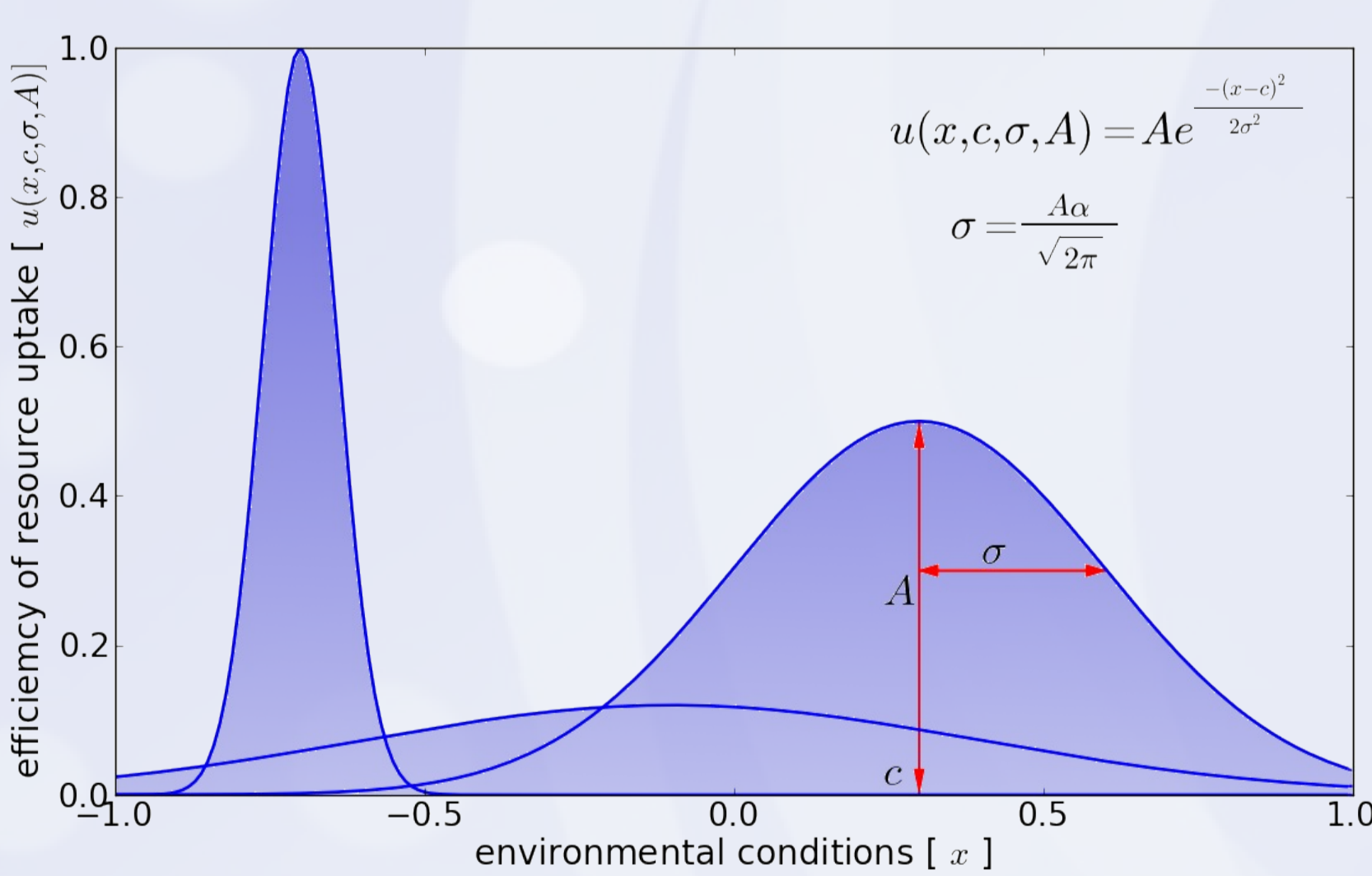
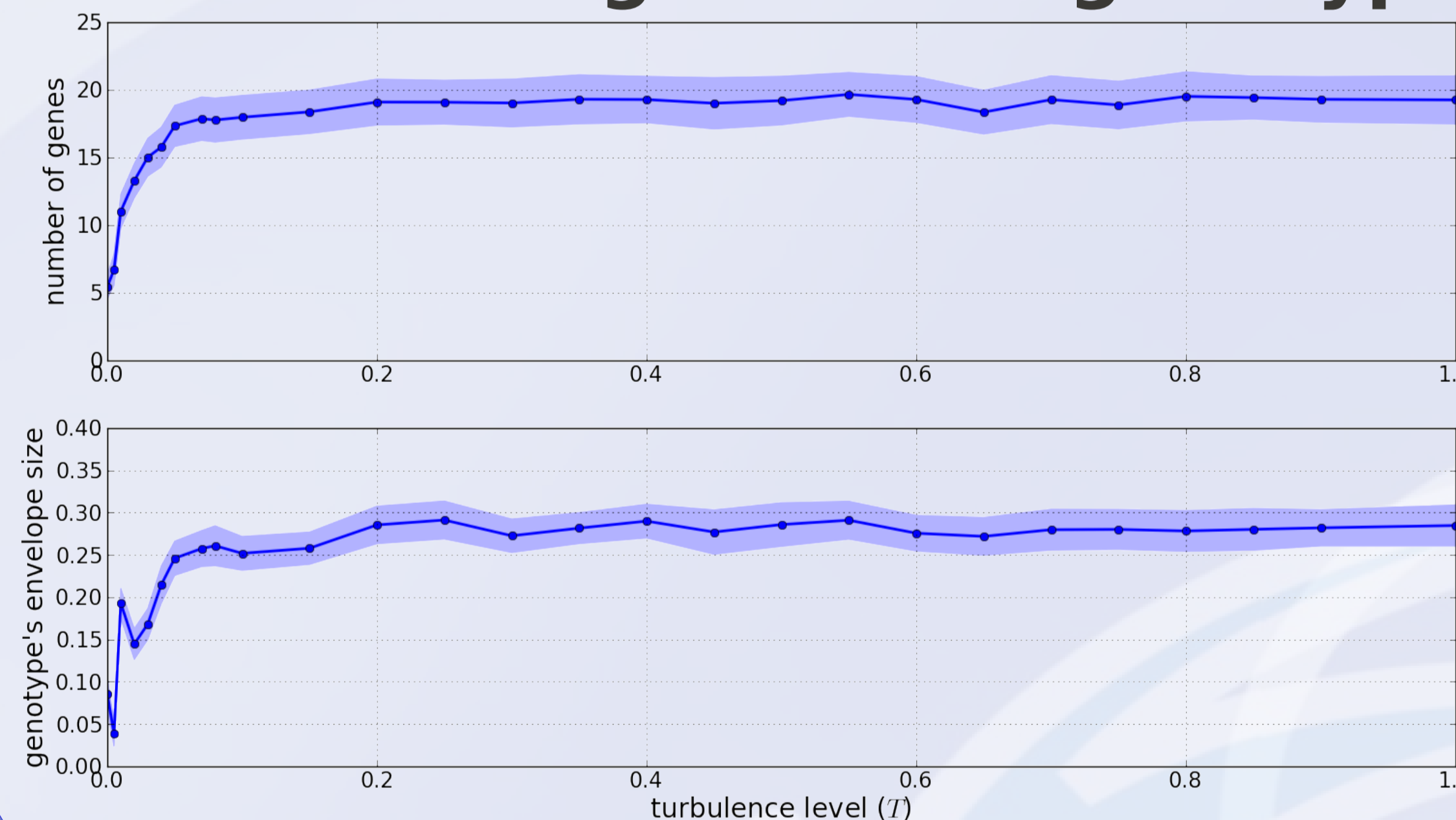


Fig.4. Example of a three-gene genotype in environment's space. Shaded area represents surface under genotype's envelope and defines cell's efficiency of resource uptake for all possible environmental conditions.

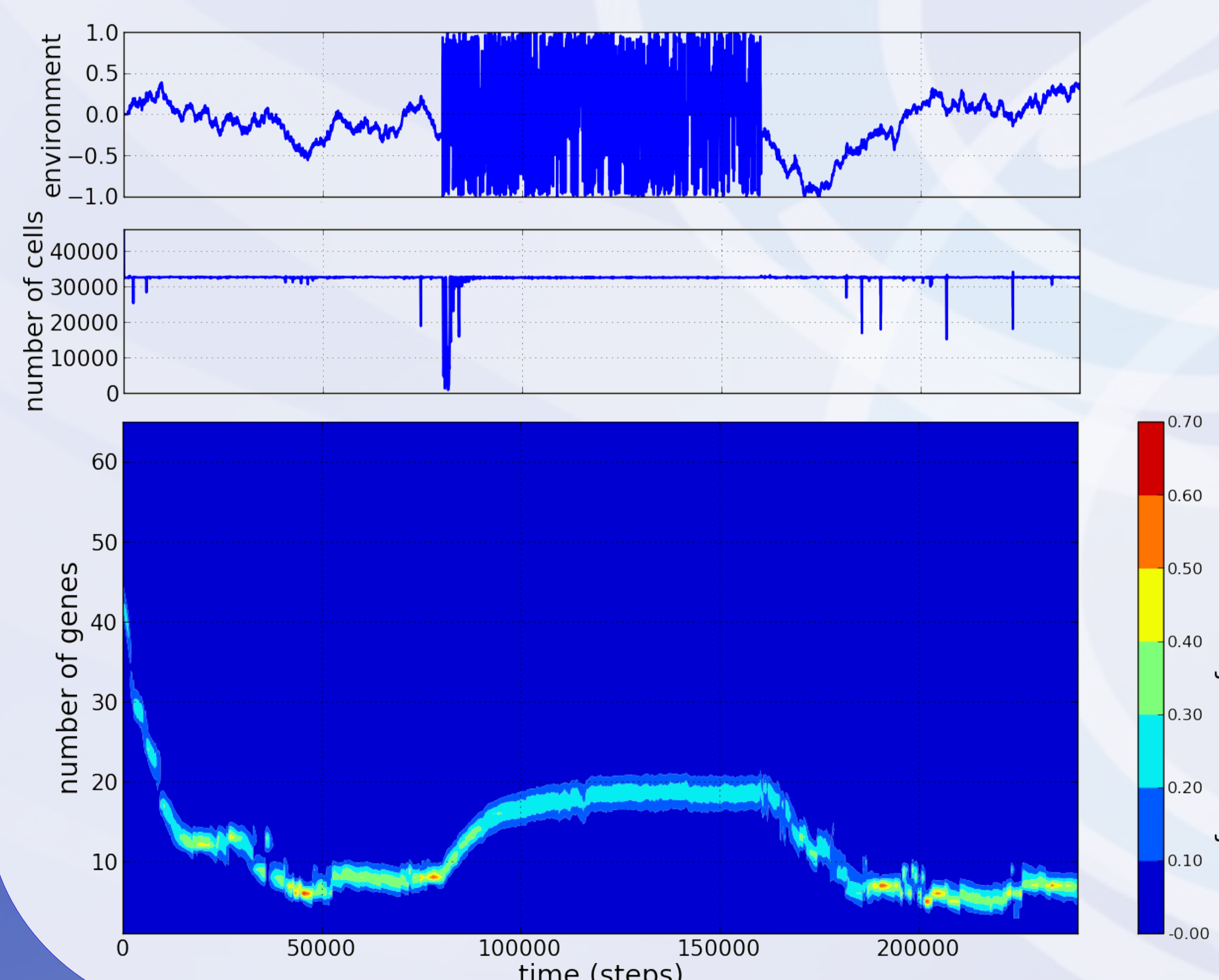
Number of genes and genotype's envelope shape



There is an optimal size of genome and an optimal shape of a genotype's envelope for a large range of turbulence levels.

Fig.5. Mean number of genes (upper panel) and mean ratio of surface under the genome's envelope to total surface of the environment (lower panel). Shaded area is standard deviation.

Response to dynamic change in turbulence levels



Number of genes can dynamically change in response to regime of environmental turbulence.

Population with larger genomes is established from small-genome population with an extinction event followed by radiation.

Fig.6. Evolution of distribution of different genome sizes in time with dynamically varying turbulence level. Initial genome size was randomly chosen between 40 and 60 genes. Turbulence level changes are shown in the upper panel.

Further development

Including lateral gene transfer in the evolution of genome structure.

Including more resources in the model.