The Amoeba-Flagellate Transformation

Camille Stephan-Otto Attolini

Institute for Theoretical Chemistry and Structural Biology,

Vienna University, Austria

Bled, Slovenia. March, 2005

The Amoeba-Flagellate Transformation

In cooperation with:

Christoph Flamm, Vienna University

Peter Stadler, University of Leipzig

Talk structure

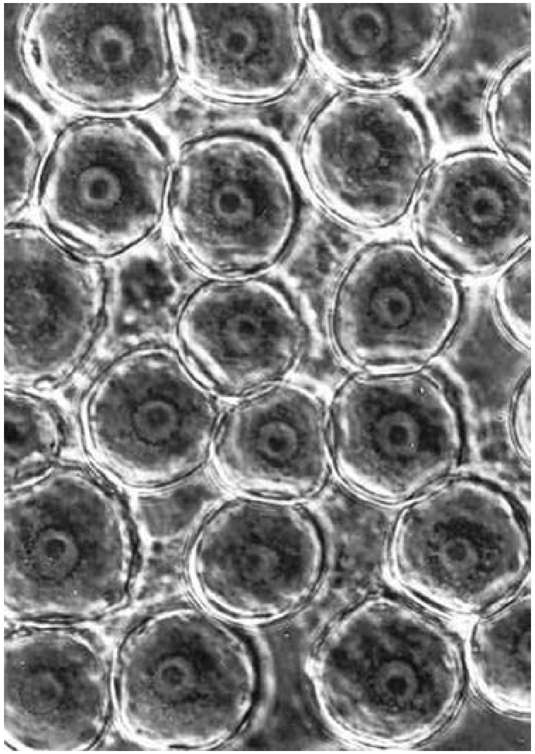
- Introduction
- The Potts Model
- The genome
- Simulation characteristics
- Results
- Future work

Introduction

- Flagellated amoebas
- Genetic Regulation

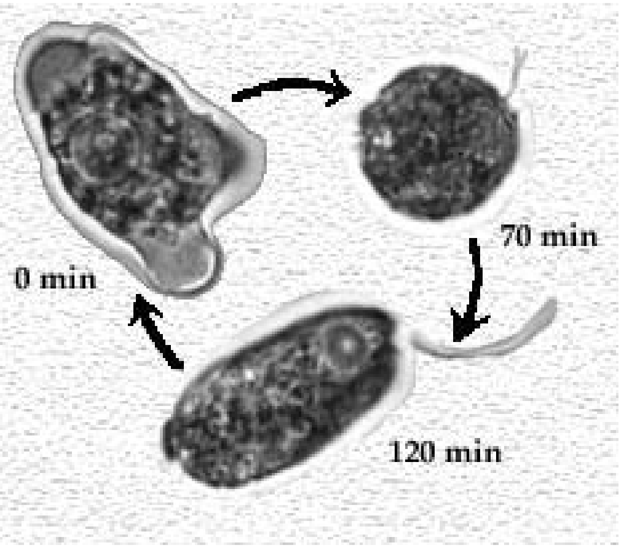
Flagellated amoebas?

The *Naegleria gruberi* is an amoeba capable of changing its structure, form and biochemistry during a lifetime reacting to changes in the environment.
Depending on the external conditions, it can transform itself from an amoeba feeding on bacteria to a flagellated cell moving in search of food.



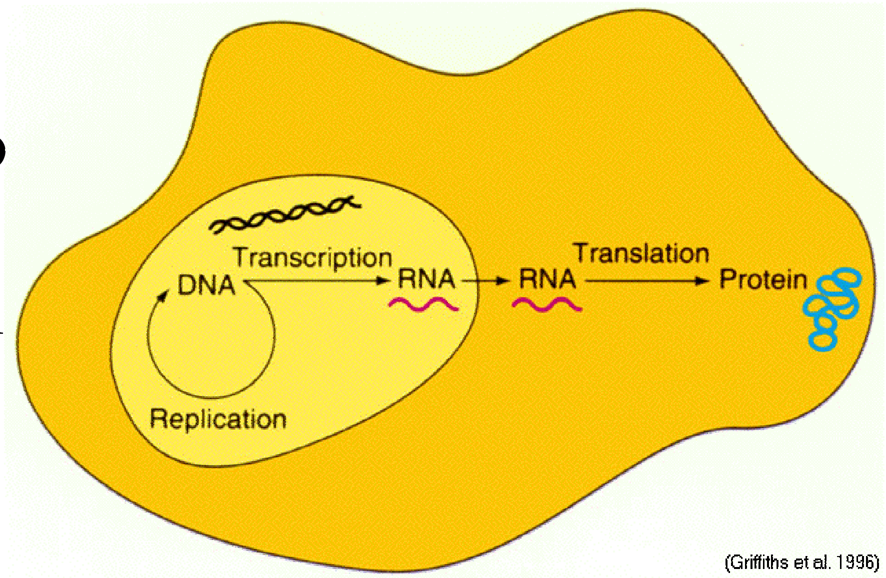
Response to the environment

When the concentration of bacteria falls, the nucleus of the *Naegleria* reacts synthesizing the mRNA for the tubulin needed for the flagella. This is a clear example of the importance of gene regulation in an organism.



Gene expression

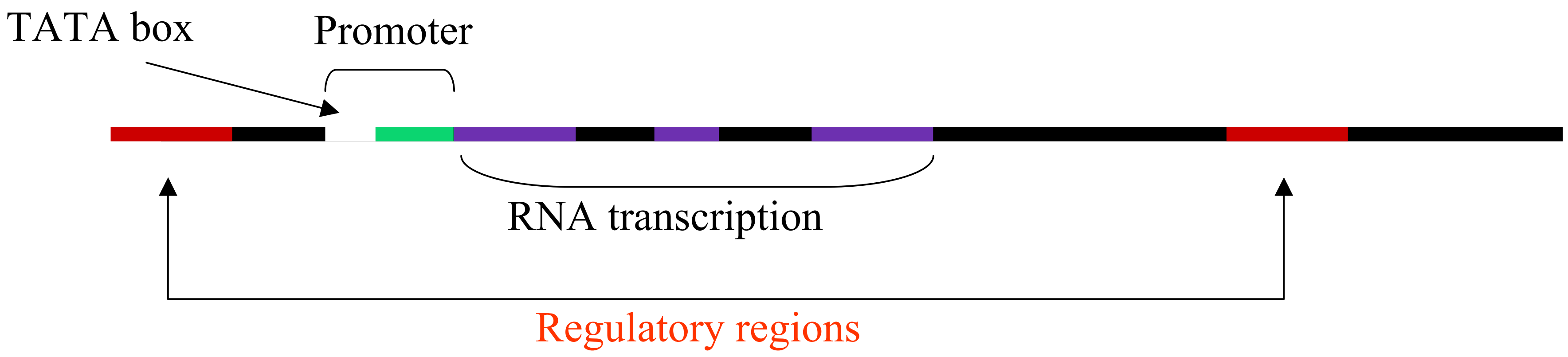
The expression of a gene is the process by which DNA is read and expressed in order to produce a functional protein. The control of this expression is performed primary at the level of DNA transcription into RNA.



Gene regulation

The regulation of a gene expression depends on multiple factors. A typical Eukaryote gene has the following structure:

- Coding region
- Promoter
- Regulatory sequences

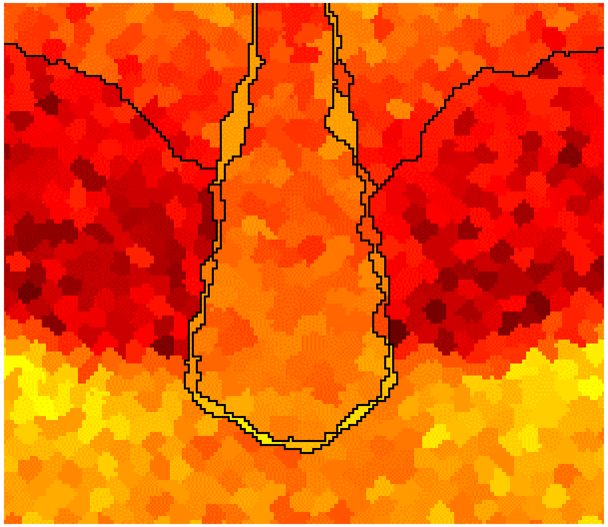


The Potts model

- Characteristics
- Minimizing energy
- Gradient sensibility

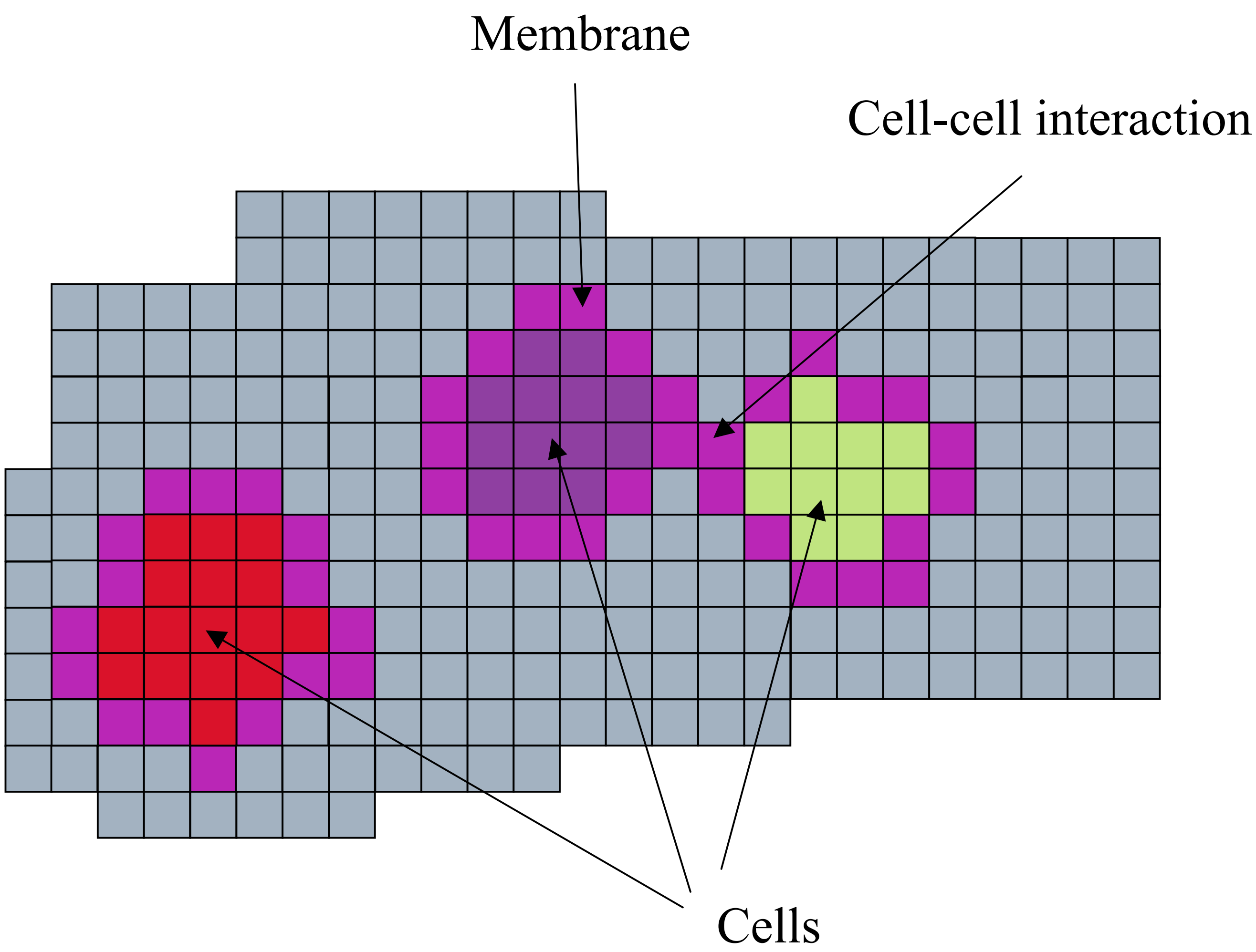
The Potts model was conceived for the first time in order to simulate and study in an effective and precise way the movement and behavior of living cells.

The model is an extension of a cellular automata with an unlimited number of states for each entry.



Dictyostelium discoideum.
A.F.M Marée

Characteristics



Cellos

Characteristics

Cells are sets of lattice points. The movement of the cell is controlled minimizing an energy function which depends on the contact of the membrane with air, substrate or other cells:

$$E_{cell} = \sum \frac{J_{type,type}}{2} + \sum J_{type,A} + \sum J_{type,S} + \lambda (v - V)^2$$

Over all entries of the membrane, where:
*-**J** is the interaction matrix between type of cells, cells and the air and cells and a substrate.*
*-**V** is the target volume and **v** the actual volume of a cell.*
*-**λ** is the compressibility of the cell.*

Cellos

The movement

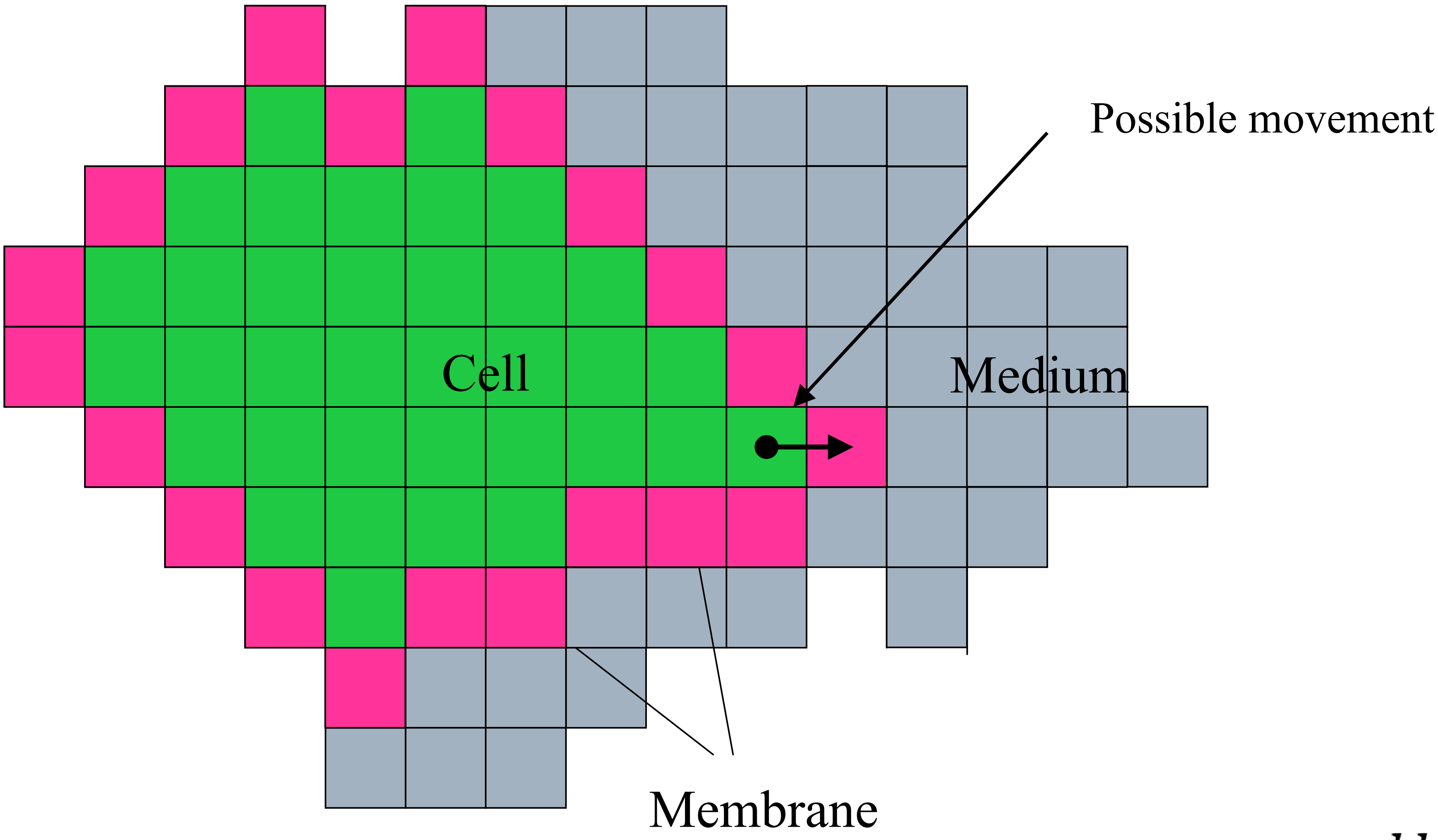
In each time step, all entries from a cell membrane are randomly chosen and copied to a neighboring position. If this change minimizes the total energy of the cell, the movement is allowed, otherwise, the probability of this change is calculated as:

$$p(m) = \begin{cases} 1 & \text{if} \quad \Delta H < -H_{diss} \\ e^{-\left(\frac{\Delta H + H_{diss}}{T}\right)} & \text{if} \quad \Delta H \geq -H_{diss} \end{cases}$$

Where $\boldsymbol{H_{diss}}$ *is the dissipation cost for the transformation and* $\boldsymbol{T}$ *is the temperature*

CellOS

The movement 2



Gradient sensibility

A first extension to the original model was to give sensibility to gradients to the cells. In order to do this, each movement following a gradient (positive or negative) is energetically improved.
The change in energy is then modified to:

$$\Delta H' = \Delta H - \mu(c_{actual} - c_{neighbor})$$

*Where **c** is the substrate concentration in each entry and*
***μ** the reaction of the cell to the substrate*

Cellos

The genome in our model

- RNA, TATA box, gene length
- Secondary structures
- Gene types
- Regulatory network

RNA and genes

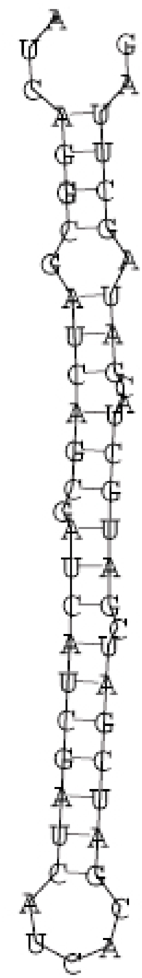
Each cell has a genome represented by an RNA string. To define genes we use a starter sequence called TATA box. In our model this string is "*GC*" and the length of genes is 40 bases. The genome is 300 to 500 bases long and is fixed during the whole simulation.

UAGCAUCAGGCGAUCAGCGAUCAUCGAUCAUCACGAUCGAUCGAUGCUACGAUAGCUUAGCGGAAC

TATA box

Coding region

Secondary structures



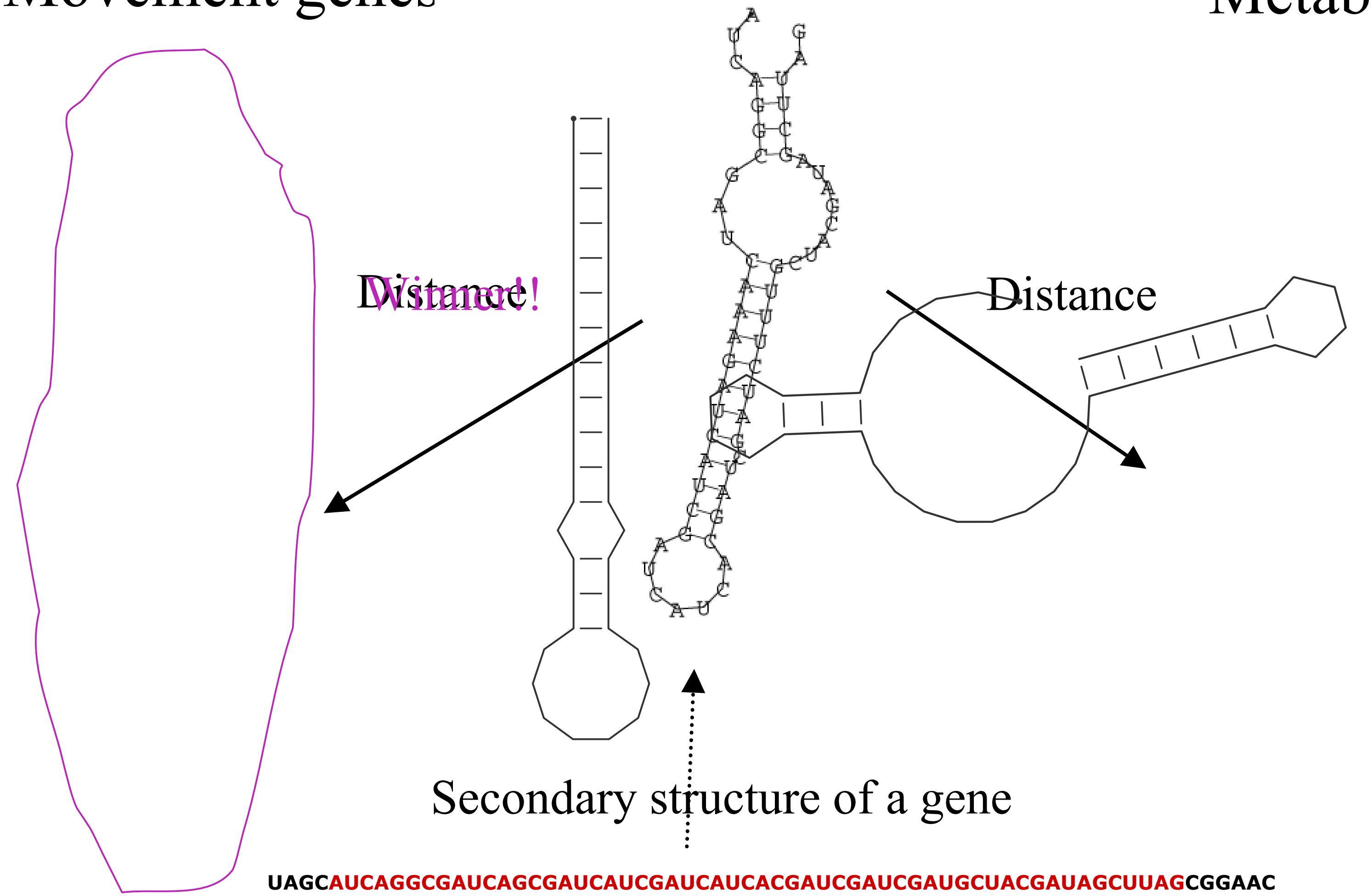
RNA molecules have the characteristic of folding into very complex three dimensional structures. Because of the lack of a fast predicting algorithm for these, it is usually used the bi-dimensional structure as a representation of the molecule's function.

UAGCAUCAGGCGAUCAGCGAUCAUCGAUCAUCACGAUCGAUCGAUGCUACGAUAGCUUAGCGGAAC

Kind of genes

Movement genes

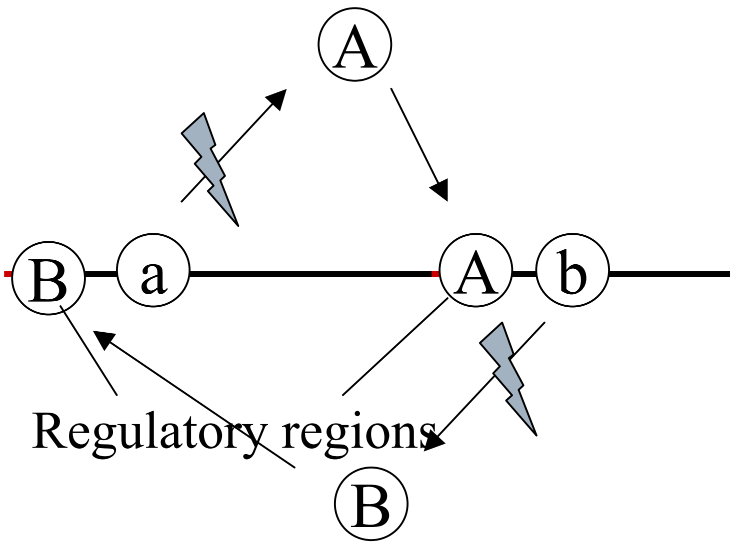
Metabolism genes



UAGCAUCAGGCGAUCAGCGAUCAUCGAUCAUCACGAUCGAUCGAUGCUACGAUAGCUUAG**CGGAAC**

Our regulatory network

At the moment, we are using a very simple regulation network for our two kind of genes:



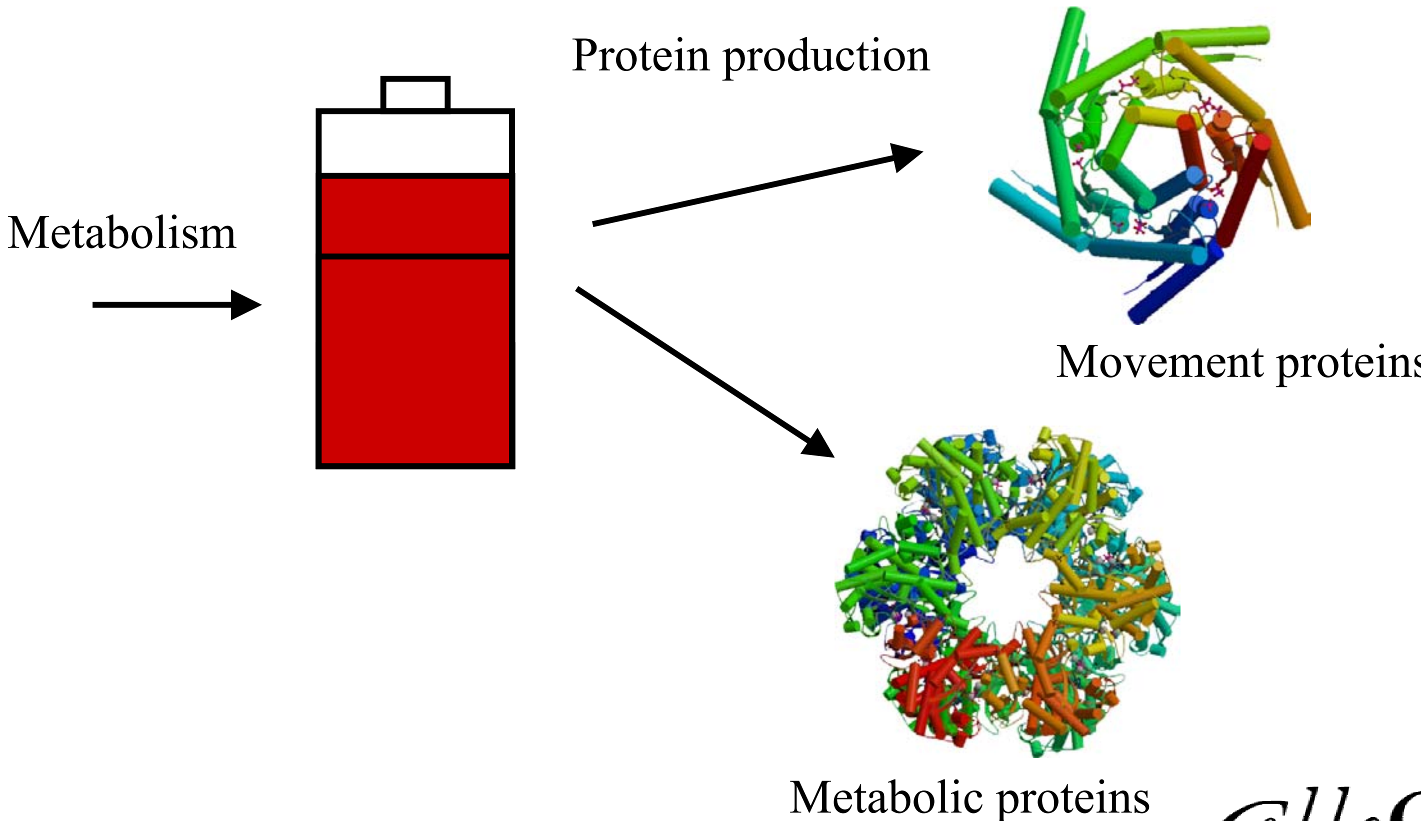
$$\frac{\partial p_m}{\partial t} = G_m \cdot k \frac{1}{1 + p_f{}^3} - d \cdot p_m$$

$$\frac{\partial p_f}{\partial t} = G_f \cdot k \frac{1}{1 + p_m{}^3} - d \cdot p_f$$

The differential equations are approximated by a Runge Kutta method of fourth order.

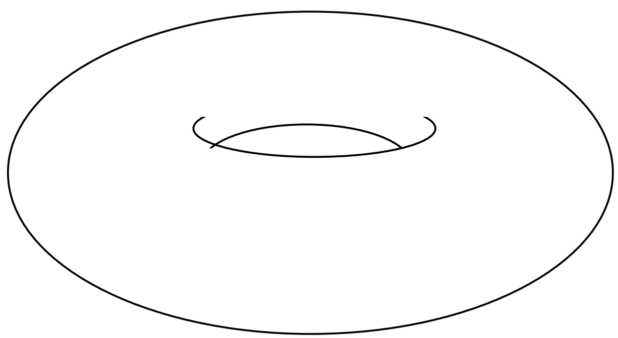
Battery

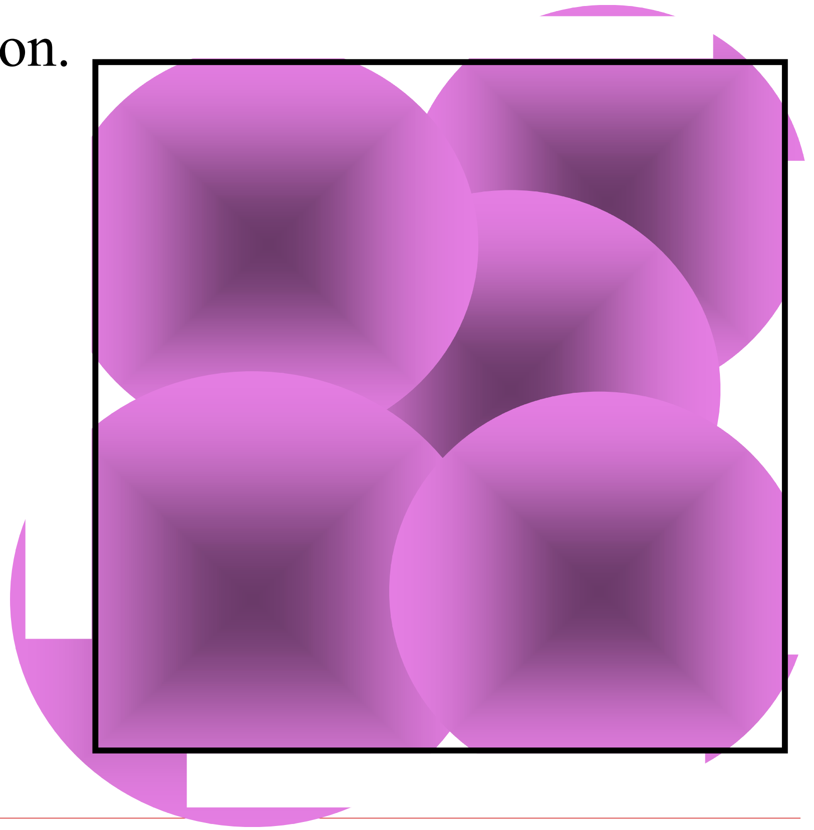
The battery is recharged every time a cell is sitting on a food spot depending on the concentration of metabolic proteins. On the other hand, battery is consumed proportionally to the production of both kinds of proteins.



Simulation basics

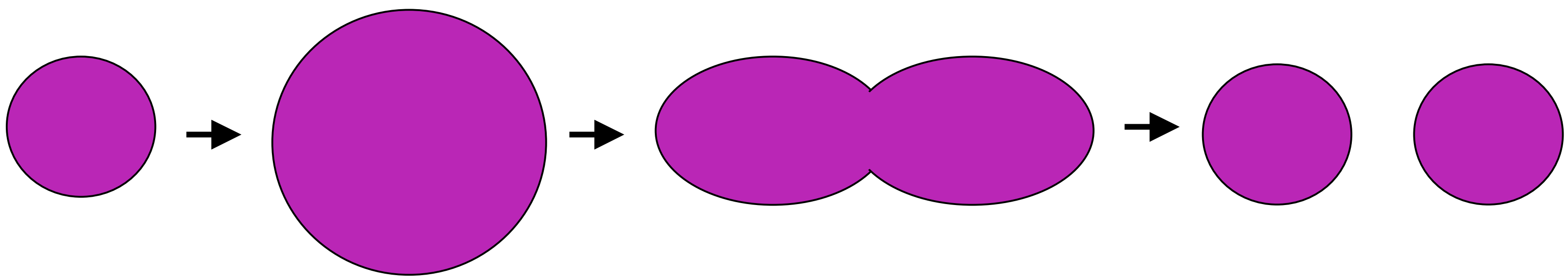
Cells live in a two dimensional lattice with the topology of a torus. Food spots are randomly distributed in the lattice, if one is depleted, a new one is created in some other point of the lattice. Cells are born with a life time defined via a Poisson distribution around a fixed parameter from the simulation.





Simulation basics 2

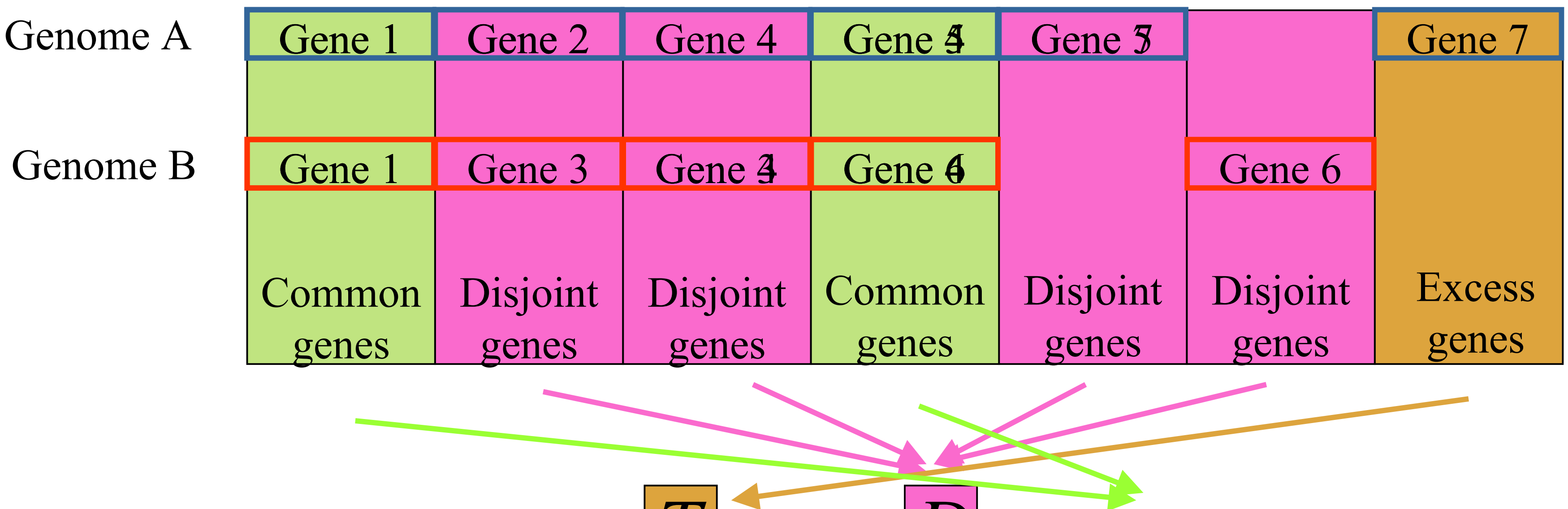
If cells are in a food spot and have the necessary metabolism genes, they are allowed to fill their battery and grow until they are double the original size and divide by fission.



In each replication a point mutation is randomly introduced in the genome.

Life time and species definition

Each gene in the population has an historical number, which we use to compare genomes and decide weather two cells belong to the same species or not.



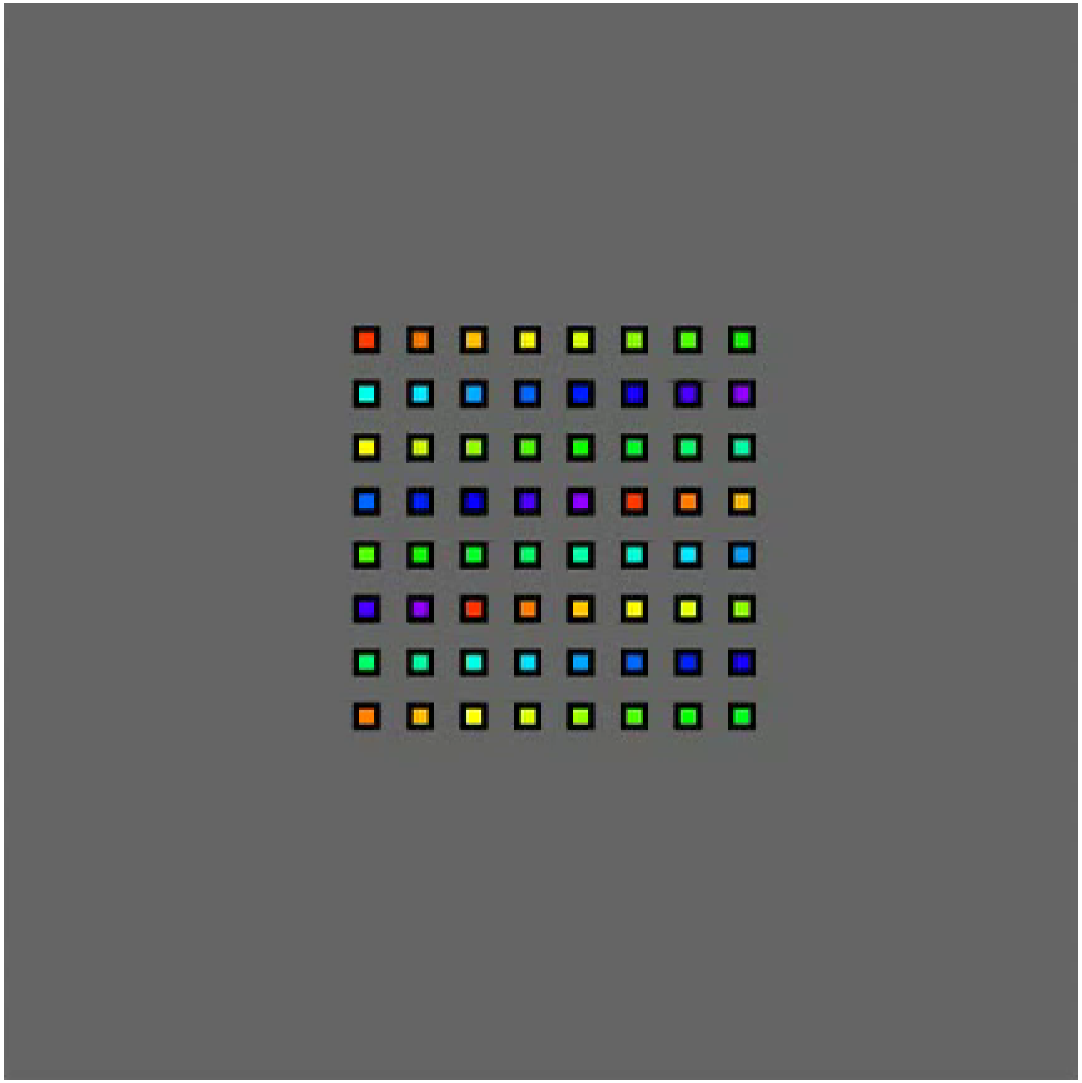
$$\delta = c_1 \cdot \frac{T}{N} + c_2 \cdot \frac{D}{N} + c_3 \cdot W$$

Results (few but nice)

- Movement genes vs. metabolism genes
- Evolution and natural selection
- Speed and lose of movement

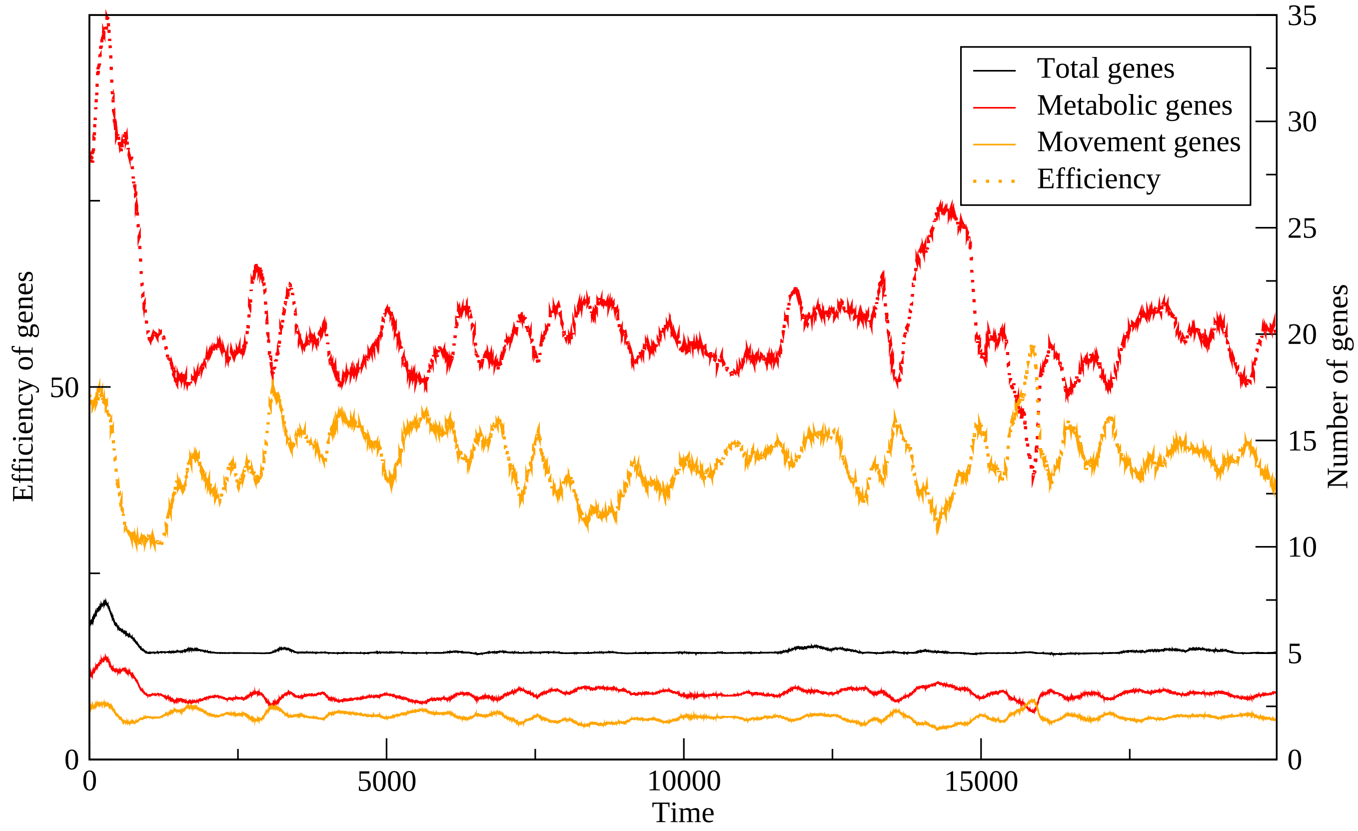
Move girls!!

Cool!

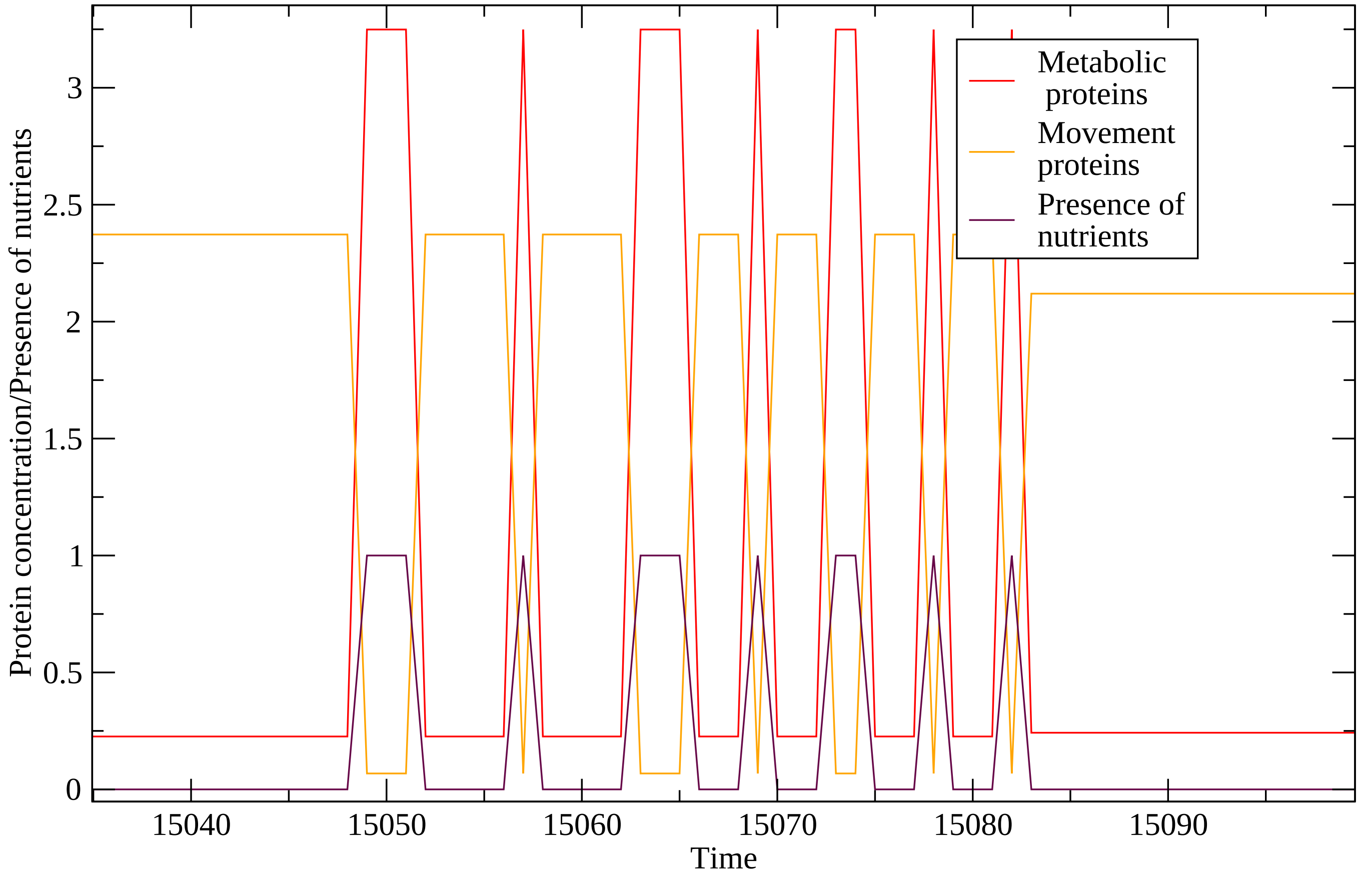


Movement vs. metabolism

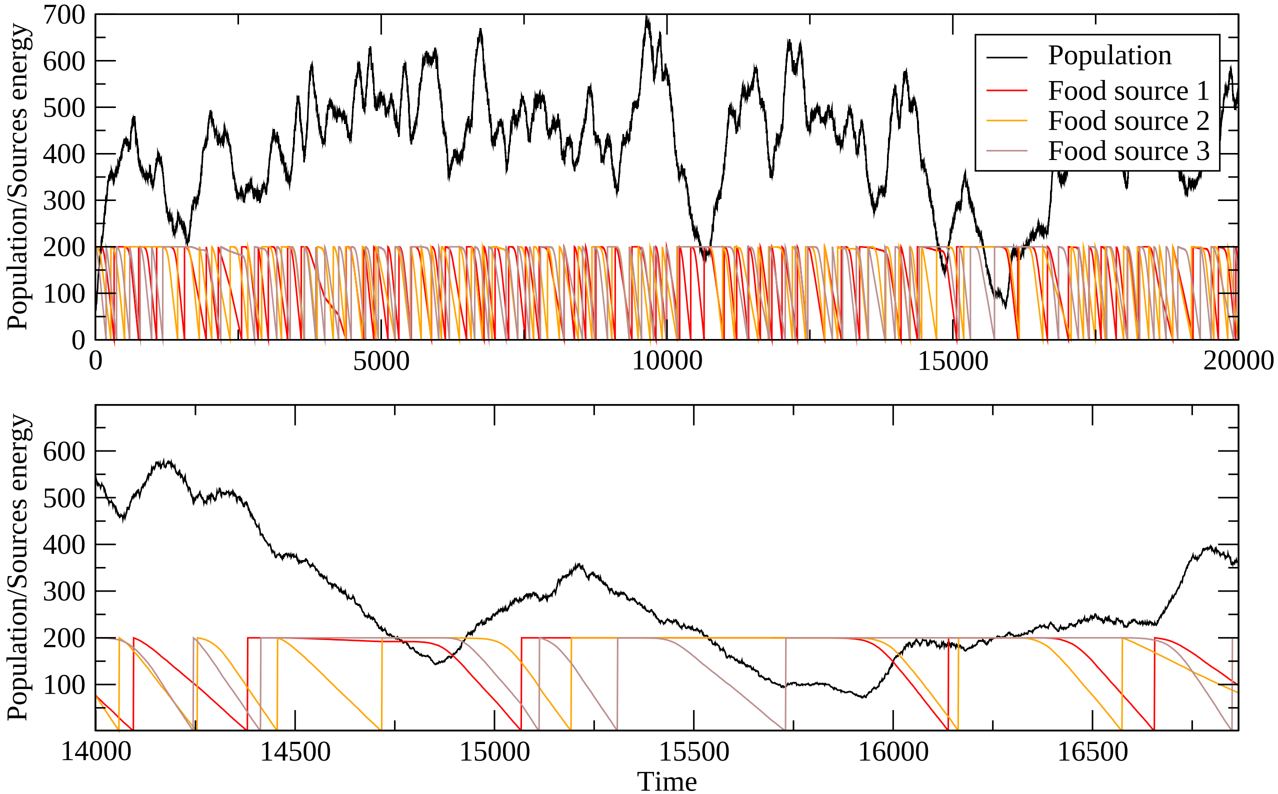
After some time, the environmen puts pressur on the genome changing th number and efficiency o genes.



Protein expression curves



Population and food sources

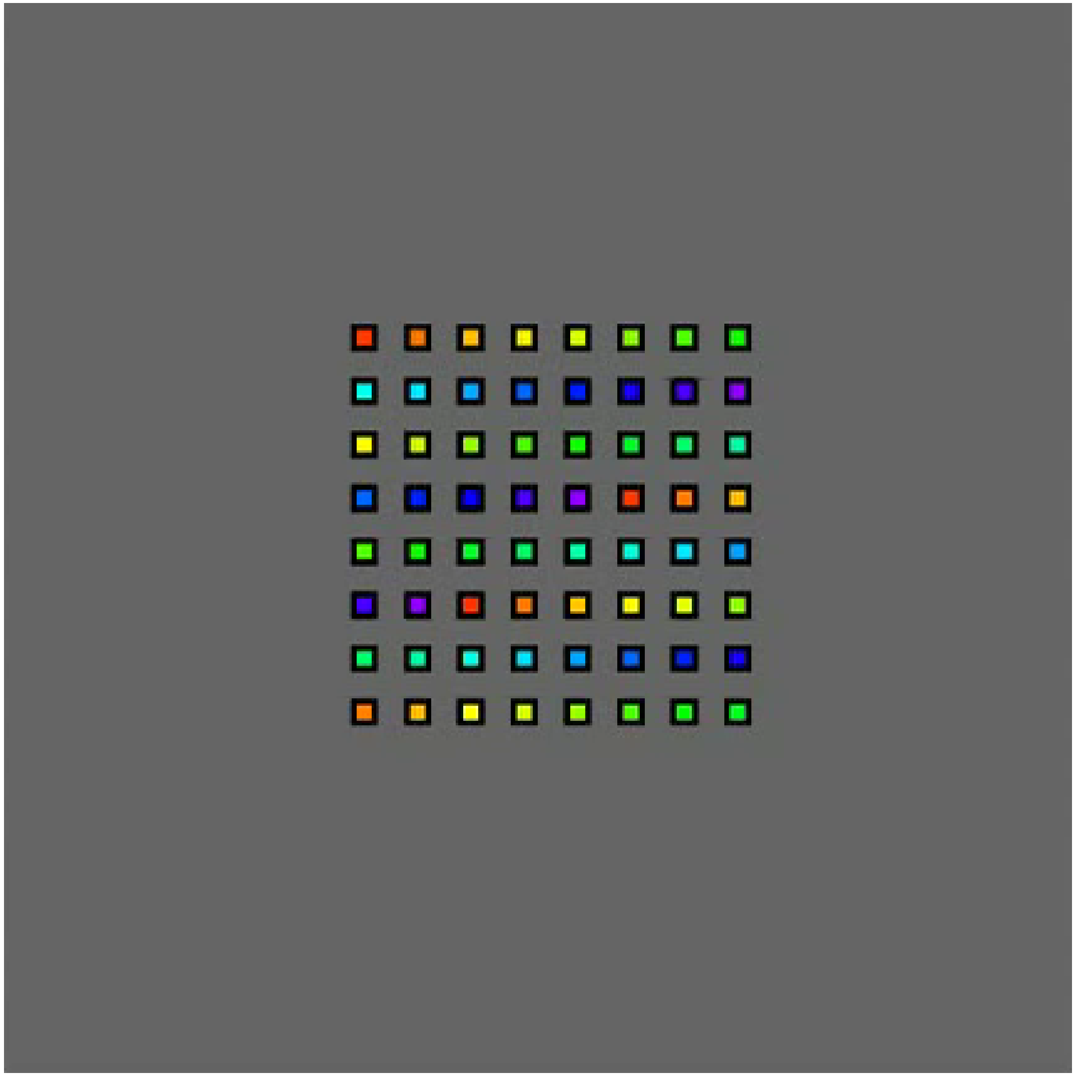


pulation grows
ıen cells are
ing from the
ırces.
pending on the
anges in the
vironment, the
mber of cells
ı fall even
ıking the
stem to
llapse.

Cellos

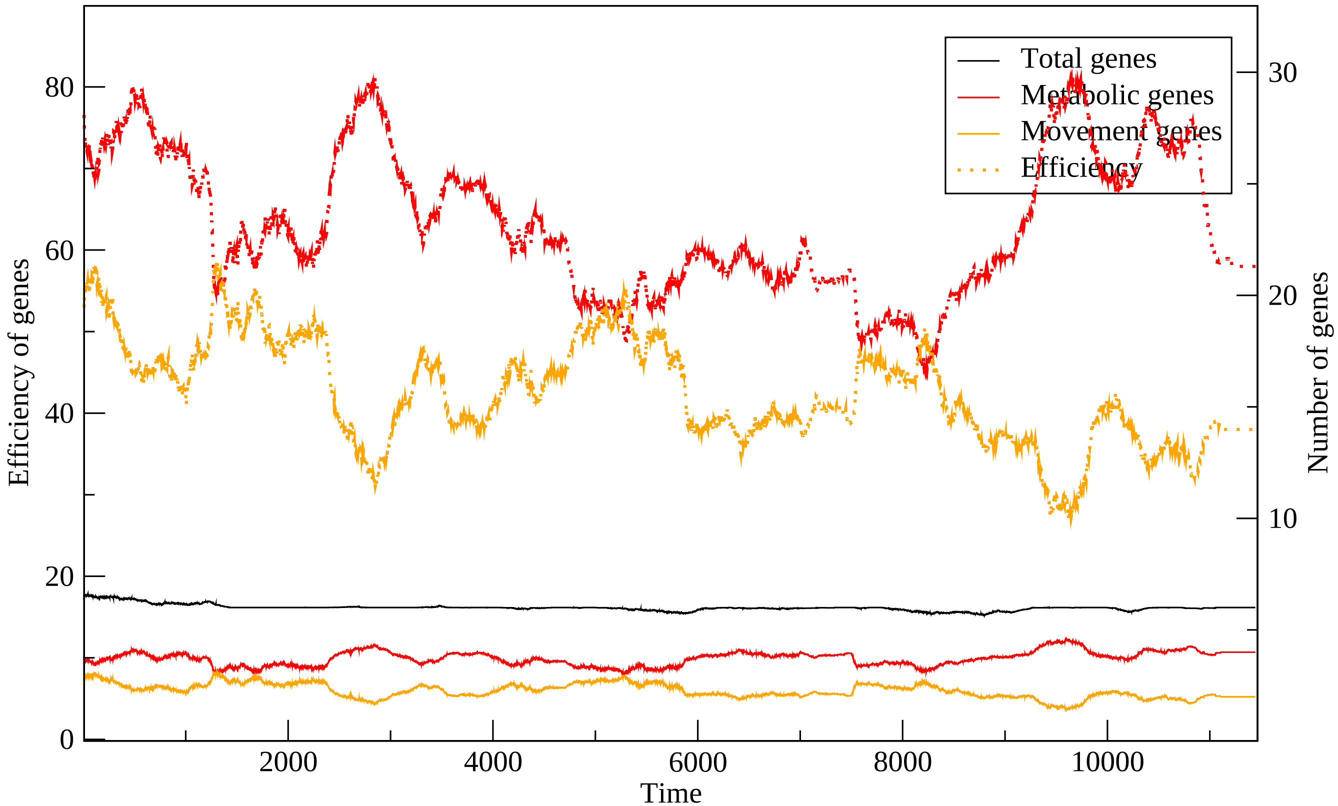
Collapse

Kill'em all

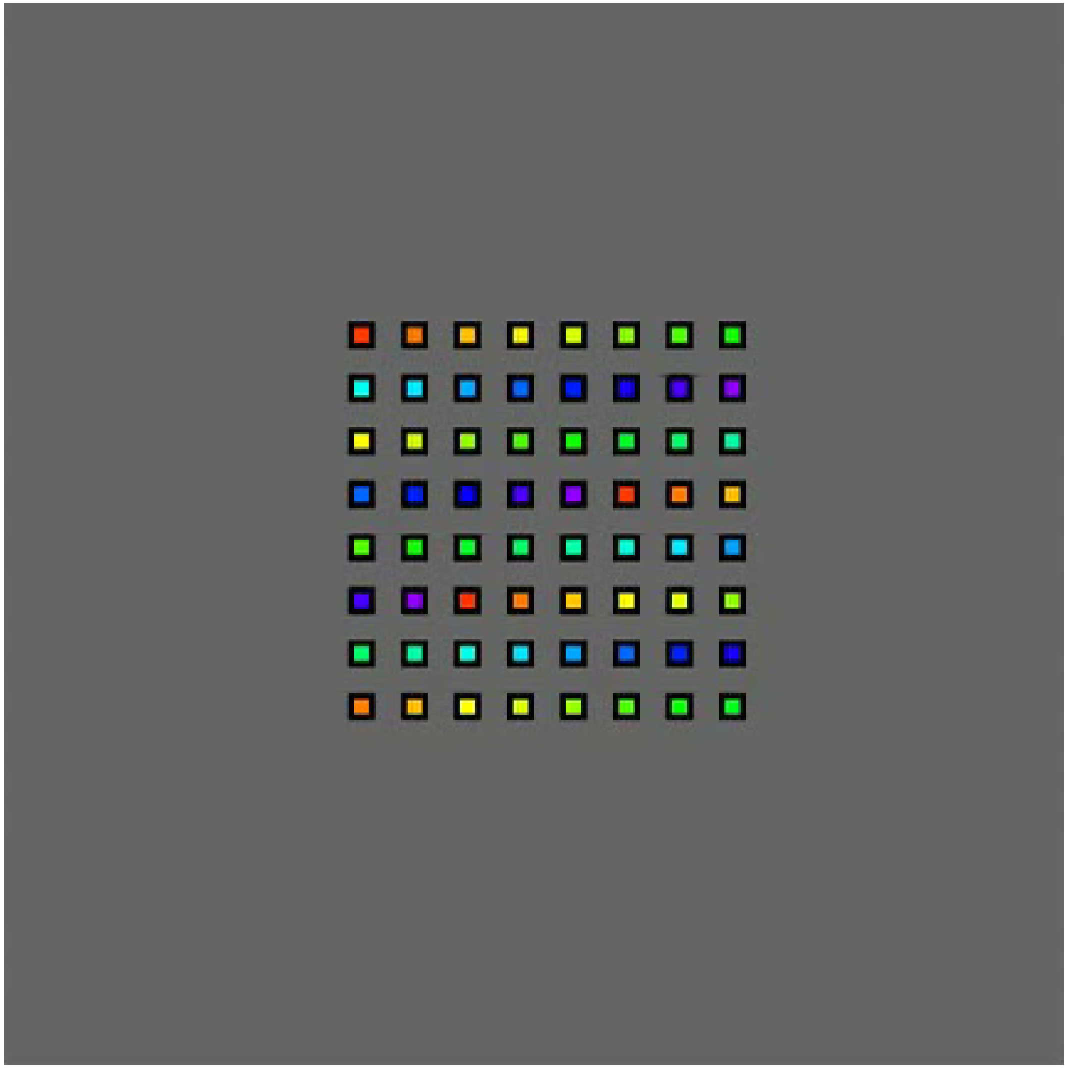


Collapse

Changes in the environment can make the system collapse. After periods of calm, a sudden change may kill all cells.



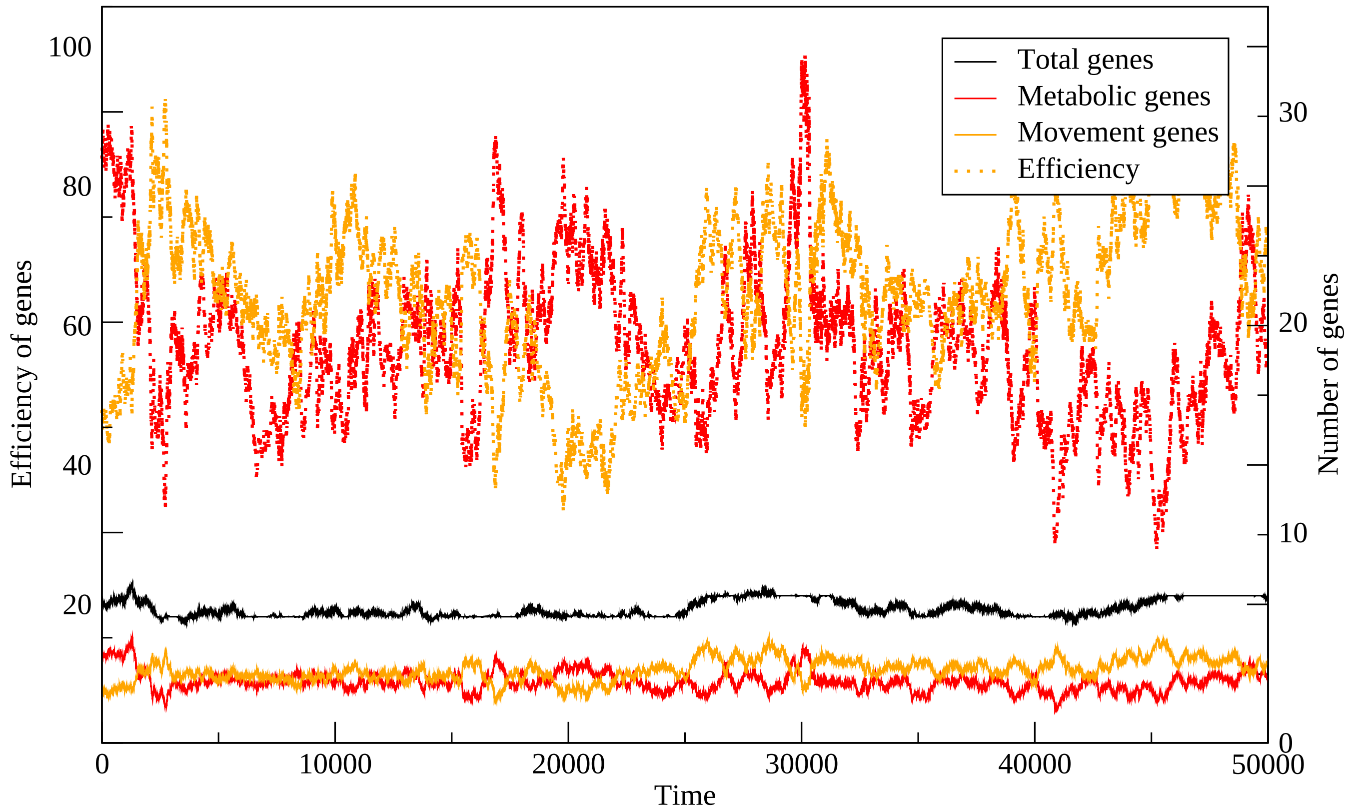
Eternal sunshine of the SPOTless mind



i.e. only one spot with eternal energy

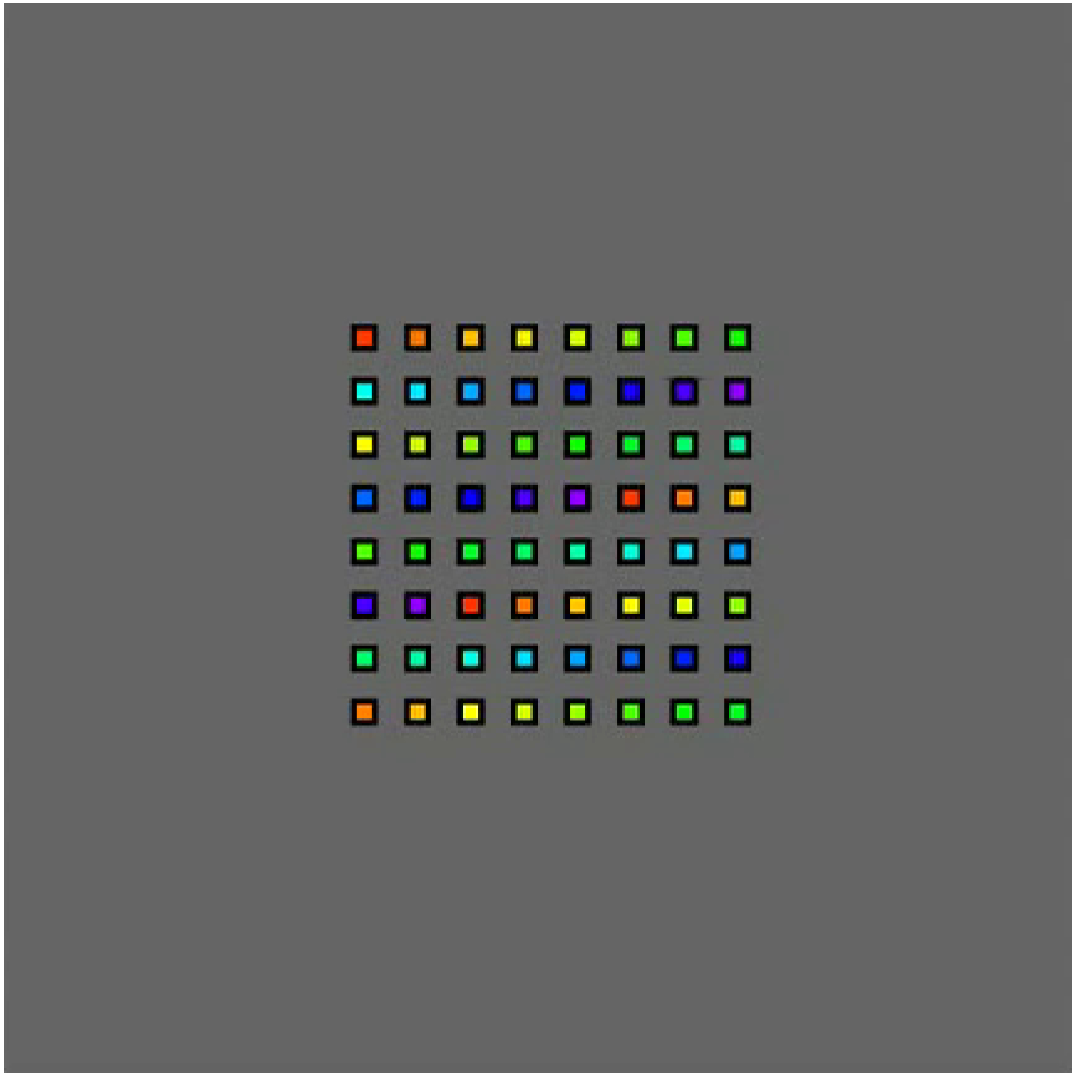
Only one spot

With only one food spot, metabolic genes may reduce in efficiency and movement genes increase until reaching the metabolic ones.



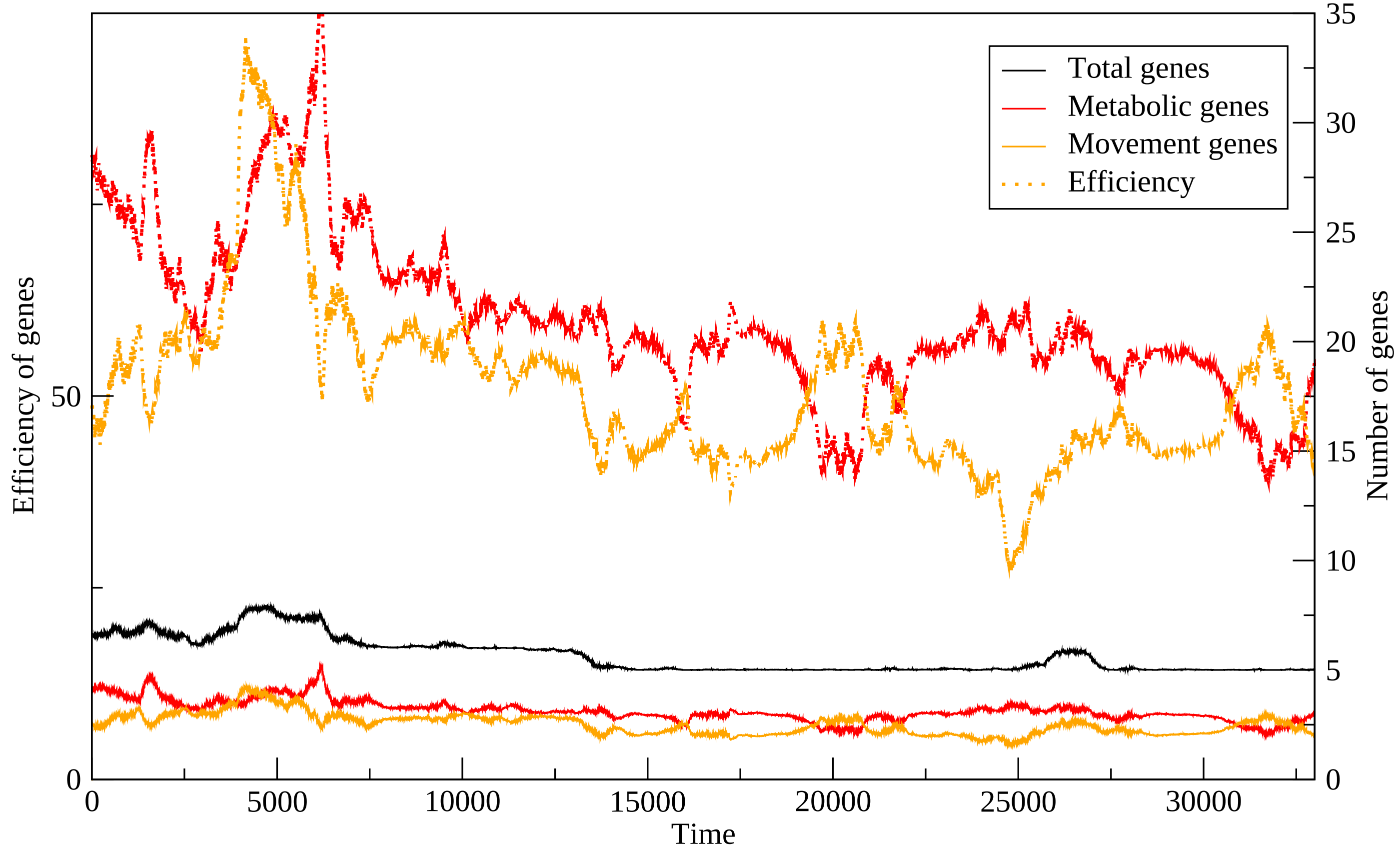
Aren’t they cute!!

Woooa!!



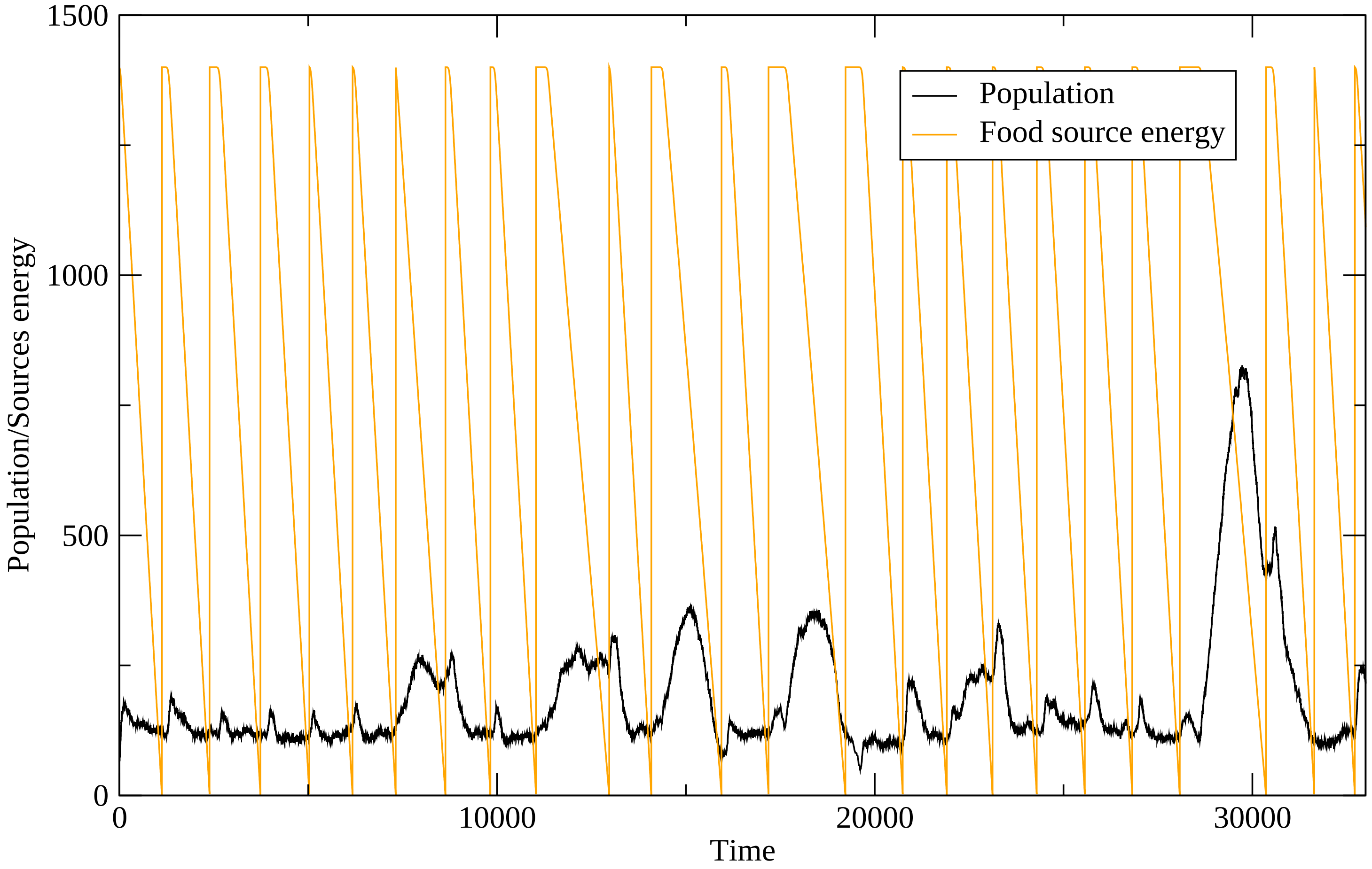
Moving food

In this run o
one food
source is in t
lattice with
energy much
longer than t
average life
the cells.



Cellos

Moving food



1500
1000
500
0
0
10000
20000
30000
Time
Population/Sources energy
Population
Food source energy

What we want to do:

- Regulation networks derived from the genome which are able to reproduce the behavior of switching amoebas.
- Investigate the phylogenetic properties when sexual and asexual reproduction is combined.

Thank you