EvoEvo Project
FP7-ICT FET Proactive EVLIT program
Project reference: 610427

Deliverable 2.3
Specifications of the population model
Version 1.7

EvoEvo Deliverable 2.3

Specifications of the population model

Due date: M6 (June 2014)

Person in charge: Guillaume Beslon

Partner in charge: INRIA

Workpackage: WP2 (Development of an integrated modelling platform)

Deliverable description: Specifications of the population model: Description of the modelling choices for the population model. This model should include a realistic population structure enabling niche construction, inter-individual interactions and open-endedness.

Revisions:

Revision no.	Revision description	Date	Person in charge
1.0	First version of the specification of the population model	11/06/14	C. Rocabert (INRIA)
1.1	Corrections by C. Rocabert and C. Knibbe	19/06/14	C. Rocabert (INRIA)
1.2	Corrections and complements by G. Beslon	27/06/14	G. Beslon (INRIA)
1.3	Corrections by G. Beslon and C. Rocabert	04/07/14	G. Beslon (INRIA)
1.4	Corrections by C. Knibbe	08/07/14	C. Knibbe (INRIA)
1.5	Validation by Paulien Hogeweg	11/07/14	P. Hogeweg (UU)
1.6	Adding a conclusion and corrections	12/12/14	G. Beslon (INRIA)
1.7	Corrections by S. Stepney	16/12/14	S. Stepney (UoY)

EvoEvo Project
FP7-ICT FET Proactive EVLIT program
Project reference: 610427

Deliverable 2.3
Specifications of the population model
Version 1.7

Table of Contents

EvoEvo Project
FP7-ICT FET Proactive EVLIT program
Project reference: 610427

Deliverable 2.3
Specifications of the population model
Version 1.7

1. Introduction

« *Some rain forests in the Amazon region occur on white-sand soils. In these locations, the physical environment consists of clean white sand, air, falling water, and sunlight. Embedded within this relatively simple physical context, we find one of the most complex ecosystems on earth, containing hundreds of thousands of species. These species do not represent hundreds of thousands of adaptations to the physical environment. Most of adaptations of these species are to the other living organisms. The forest creates its own environment.* »

Thomas S. Ray (1995)
Picture: detail of *Surprised!*, Henri Rousseau (1891)

In an evolutionary model, the environment is a crucial element. It imposes external constraints on the evolving organisms, eventually resulting in the selection pressure due to the limited carrying capacity. Moreover, the organisms living in an environment also modify it. This creates new evolutionary niches in which new species can emerge, eventually creating a complex ecosystem and a complex trophic network. Most preexisting models used as a baseline for WP2 tasks do not integrate an evolvable environment. For example, the environment in aevol (Knibbe et al., 2007) is a mathematical function $E: [0,1] \rightarrow [0,1]$, identical for the whole population even if it can vary over time. Resource cycling has been observed with the "Pearls-on-a-String" model (Crombach & Hogeweg, 2009), but here too, environment was strongly simplified (metabolites being input and output binary strings of the genetic regulation network of each individual).

The objective of task T2.3 is to use the knowhow of Inria and Utrecht University (and others, e.g. the Tierra model, see Ray, 1991) to propose an integrated environmental model able to drive the emergence of a complex ecosystem. This model will in particular include temporal variations (either cyclic or random), random noise, and metabolites release/diffusion/consumption. It will thus enable us to study how the population structure contributes to the "evolution of evolution" phenomenon. Ultimately, the objective of the population model is to enable open-ended evolution in the integrated model. Open-endedness is here defined as the ability of the evolution to permanently create new challenges due to changes induced by evolution itself. These changes may occur at different levels (molecules, cell, populations…). Therefore open-endedness can emerge at these various levels. In this project, we choose to study it mainly at the population level since it is likely to be easier to implement than open-endedness at the molecular and cellular levels.

One of the most important components of open-endedness is *interactions*: For an evolutionary change occurring on an entity A to modify the evolutionary challenges of an entity B, A and B must interact and this interaction must result in a drift of B's target and/or in the emergence of new ways to reach/perform this target and/or in the emergence of new targets (possibly leading to radiation). This means that this interaction must be more complex than the competition for reproduction (which is generally the sole inter-individual interaction between individuals in evolutionary computation and even in most artificial life setups). That is why the main objective of the population model proposed here is to define the way individuals will interact. To this aim, this document will be organized in three parts. First, we will summarize the main choices made in deliverable D2.1

EvoEvo Project
FP7-ICT FET Proactive EVLIT program
Project reference: 610427

Deliverable 2.3
Specifications of the population model
Version 1.7

(specifications of the genome-network model) as they will be the base on which we will define the inter-individual interactions. Then we will describe the environment of the individuals, which will serve as a mediator for inter-individual interactions. Finally, we will describe the new features of the individuals that will be necessary to allow for efficient interactions.

2. Specifications of the population model

2.1. Reminder of the artificial chemistry described in deliverable 2.1

Each active gene in a genome codes for a tuple (s, p, K_{cat}, K_m), s and p being the substrate and the product of an enzymatic reaction ($s, p \in \mathbb{N}^*$), K_{cat} and K_m being the constants of the following Michaelis-Menten equation:

$$\frac{d[p]}{dt} = \frac{(K_{cat} \times [s])}{(K_m + [s])}$$

If the gene is expressed, a transition occurs in the metabolic space between s and p at a certain rate depending on K_{cat} and K_m. The transition can occur in the cell, or between the cell and its environment (in this case, the tuple codes for a pump). Hence, for each cell, the whole genome produces enzymes defining a metabolic network, where nodes are molecular species involved in chemical reactions, and edges are enzymatic reactions encoded in the genome (figure 1).

To sum up, we defined an artificial chemistry $\{S, R, A\}$ where:

- The set of molecules S is the integer space $\mathbb{N}^*$,

- Each cell carries its own set of reaction rules R, defined by its genome,

- The system of ordinary differential equations defining a cell's metabolic network is integrated using a continuous and deterministic method.
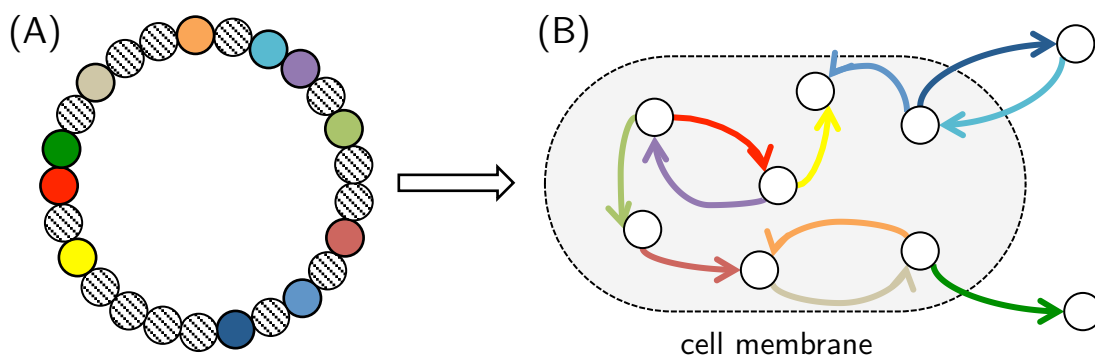

Figure 1 - The genome codes for a metabolic network. (A) Each active gene produces enzymes catalysing reactions of the form s → p, where enzyme activity is implicit. (B) The set of enzymes produced by the genome defines a metabolic network, with metabolic pathways, cycles and environmental exchanges.

EvoEvo Project
FP7-ICT FET Proactive EVLIT program
Project reference: 610427

Deliverable 2.3
Specifications of the population model
Version 1.7

2.2. Physical environment

Open-endedness requires individuals to interact with each other. However, inter-individual interactions must not occur at the whole population level (i.e. all individuals interacting with all individuals), otherwise there is a strong risk to reduce the variability between individuals. That is why, following the work of P. Hogeweg's group, we will use a spatial structure. Individuals are dispatched on a 2D lattice of size $W.H$ (with W the width and H the height of the lattice), each cell containing at most one individual. The physical environment is also described at the lattice level: each lattice cell contains a list of free metabolites, each with its concentration level. Those free metabolites diffuse and degrade (figure 2), both processes being controlled by two parameters: the diffusion rate D and the degradation rate D_g.

- Diffusion: at each time step, a proportion of the free metabolites of a given cell *i* are spread to its neighbours (we will use a Moore neighbourhood, figure 2.C). This proportion will be controlled by the diffusion parameter D.
- Degradation: at each time step, a proportion of the free metabolites of a given cell *i* are removed. This proportion is controlled by the degradation parameter D_g.

Individuals compete for the free metabolites and to produce offspring in empty cells. Individuals interact with their local environment by pumping metabolites in and out and releasing their content at death (see below).
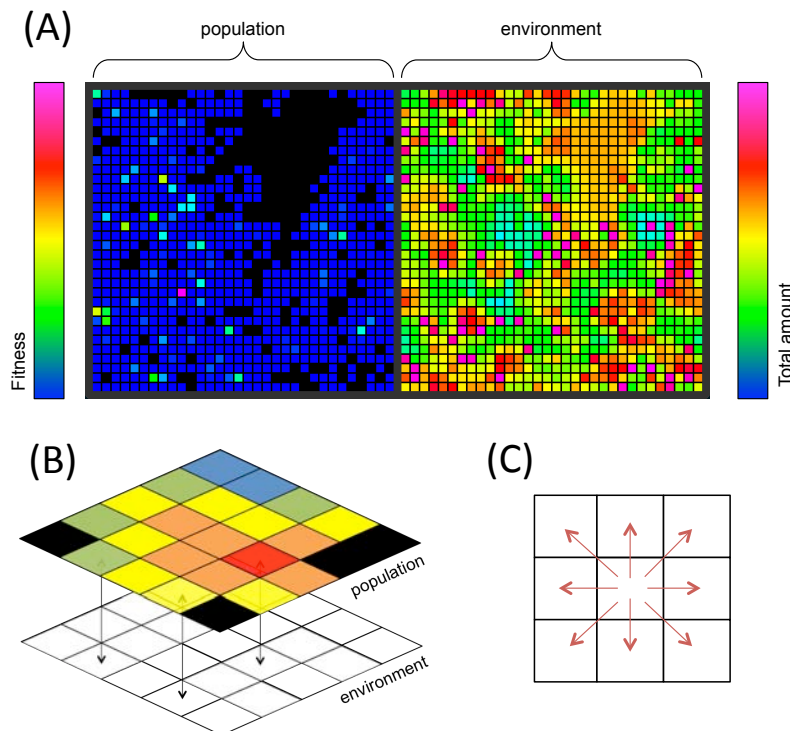

Figure 2 – Overview of the population level. (A). Population and environment share the same lattice (a tore). On the left part, individuals reproduce in gaps (black cells), and are coloured depending on their fitness. On the right part, the environment is defined by free metabolites diffusing in the lattice and degrading. (B) Each individual interacts with its local environment, by pumping in and out metabolites. (C) Free metabolites diffuse in the Moore neighbourhood (the eight cells surrounding the central one).

EvoEvo Project
FP7-ICT FET Proactive EVLIT program
Project reference: 610427

Deliverable 2.3
Specifications of the population model
Version 1.7

The formalism presented above allows for several outcomes. Firstly, population size is variable. Depending on growth rate, lattice size, and interactions between individuals, simulations can lead to steady state, but also to oscillations or even extinctions. Secondly, since individuals can interact with their surroundings via intake and release of metabolites, interactions between individuals can emerge such that necrophagy, share of public goods, arms race by releasing toxic metabolites, and so on.

In order to vary the strength of spatial structure, a random fraction mig of organisms and free metabolites can be swapped at each time step. To do so, pairs of lattice cells are randomly chosen and their contents are swapped. Depending on the mig value, one can vary the population structure from well-mixed (mig = 1) to perfectly local (mig = 0). Moreover, some metabolites can also be artificially maintained at a constant concentration, be regularly provided locally or globally in the environment or, on the opposite, be regularly washed-out from the environment. Thus, various real experimental setups can be mimicked, including serial plates or chemostat (Mozhayskiy & Tagkopoulos, 2012). Similarly, some individuals can be regularly picked up in the environment to seed a new colony, thus mimicking a mutation accumulation experiment. All these optional features will be useful for further experiments in WP3 as mutation accumulation is often used in experimental evolution (figure 3). In order for the simulation to be computationally tractable, a concentration threshold may be defined above which a given metabolite could be considered absent from the local environment.
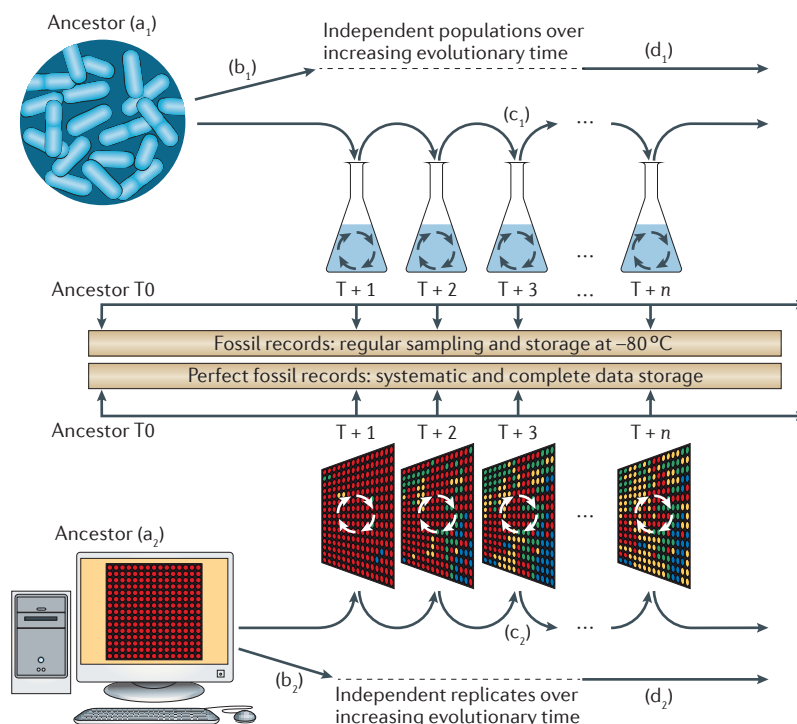

Figure 3 – From (Hindré et al., 2012). Parallel in vivo (top) and in silico (bottom) experimental evolution. The experiments conducted in vivo will be mimicked in the computational framework.

To summarize, the environment model will be characterized by the following parameters:

EvoEvo Project
FP7-ICT FET Proactive EVLIT program
Project reference: 610427

Deliverable 2.3
Specifications of the population model
Version 1.7

- Environment size $W.H$ ($W.H$ thus being the maximal carrying capacity[1] of the environment),
- Diffusion coefficient D,
- Degradation rate D_g,
- Metabolites inflow I_i (for each metabolite i),
- Metabolites outflow O_i (for each metabolite i),
- The migration coefficient mig.

Note that the dynamics of inflow/outflow can be constant (i.e. applied at each time step) or time dependent.

At a given time t each lattice cell of coordinate (x, y) is characterized by:

- The individual that occupy the cell (possibly null)
- The list of free metabolites that are available and their concentrations $C_i(t)$

Given the parameters of the environment, the dynamics of a free metabolite i in a cell is then given by:

$$C_i(t+1) = C_i(t) - DgC_i(t) + \sum_{j \ in \ neighbours} D \ C_{i,j}(t) - 8DC_i(t) + I_i(t) - O_i(t)$$

$C_{i,j}$ being the concentration of metabolite i in the lattice cell j.

2.3. Interactions between individuals

The environment model described above enables interactions between individuals through modification of a shared environment. For these interactions to be efficient and effectively lead to complex evolutionary dynamics and open-endedness, individuals must be able structure their environment and the environment of their neighbours, thus influencing their "lives" and evolutionary fate. To do so, we propose to slightly modify the individual model (as described in deliverable D2.1 and briefly summarized above). More precisely, we provide the individuals with three properties:

1. The ability to release metabolites in the environment,
2. The ability to intake metabolites from their environment,
3. The possibility to indirectly modify the inner metabolites concentration of their neighbours, due the permeability of the cell membranes.

These properties are achieved through special proteins called pumps. As explained above, in our Artificial Chemistry, genes encode enzymes that transform a substrate (actually an integer $s \in \mathbb{N}^*$) into another one, the product (another integer $p \in \mathbb{N}^*$) with a specific rate. We will consider that the enzymes for which the substrate and the product are equal (i.e. $s = p$) are inflowing or outflowing pumps for the metabolite s. The orientation of the pump (in- or outflow) and its efficacy are specified by the reaction constants encoding by the gene tuple. Using such pumps, the individuals are able to control their cytoplasm composition and to maintain their internal homeostasis.

---

[1] The actual carrying capacity may depend on the species needs and on the metabolites available in the environment. In a complex situation where the environment contains different co-evolving species, the carrying capacity of each species may depend on the metabolites released by the other ones and on the dynamic of metabolites inflow/outflow.

EvoEvo Project
FP7-ICT FET Proactive EVLIT program
Project reference: 610427

Deliverable 2.3
Specifications of the population model
Version 1.7

The pumps are important mechanisms for the cell to maintain a stable metabolic activity. However, by themselves, they do not allow for complex cell-to-cell interactions since each individual can fully and autonomously control its internal composition. That is why we added two mechanisms allowing for resource cycling and complex cell-to-cell interactions:

- "Necrophagy": each time an individual dies, all the metabolites it contains are released in the local environment and start diffusing on the grid. This mechanism will lead for a progressive complexification of the environment as long as evolution creates more and more complex individuals.
- "Permeable membrane": the cell membranes are not perfect barrier for metabolites. As well as metabolites diffuse on the environment grid, they diffuse from the grid cell (i.e. the local physical environment of the individual) to the individual itself with a diffusion coefficient D_m. At each time step, a fraction D_m of the metabolites diffuse through the cell membrane, resulting in a progressive balancing of the metabolites concentrations in the environment and in the cell. Thus, pumps are active mechanisms that the cell can use to maintain an internal concentration different from the external one[2]. Consequently, a metabolite actively released by an individual in its environment will diffuse to the neighbouring lattice cells (with diffusion coefficient D) and to the individuals that "live" there (with a diffusion coefficient D_m), possibly perturbing their internal metabolic activity, unless these individuals have evolved mechanisms to protect themselves against these perturbations.

These two additional mechanisms are likely to initiate complex dynamics at the ecosystem level (creation of a trophic network, niche construction…) and at the cell-to-cell level (release of public good or, on the opposite, of bacterial toxins).

2.4. Additional features for the population model

In order to increase the level of complexity and realism of the population level, some features could be added to the current specifications. In particular, the following properties could be added to the model for specific experimental purposes:

- Some metabolites could act as toxins by killing the bacteria that accumulate them in their cytosol. This accumulation could be the consequence of toxin production or toxin absorption. The individuals may evolve protective mechanisms by either degrading the toxins through specific enzymes or excreting them thanks to outflowing pumps.
- Only the metabolites (S, see section 2.1) can be present in the environment. We consider here that enzymes (R, see section 2.1) are rapidly degraded when released out of the cell. This assumption is coherent with most biological situations. However release and diffusion of enzymes could be included in the model for specific purposes such as the modelling of endosymbiosis, a situation in which the proteins produced by symbiotic bacteria could be transferred to its host through the environment.
- Horizontal Gene Transfer (HGT), being a mutation operator, is not including in the population model. However, some HGT mechanisms will need adaptation of the environment model to allow for:

---

[2] Depending on the D_m value, the pumps will precisely control the internal composition of the cytoplasm ($D_m = 0$) or they will have no effect ($D_m = 1$). In between, pumps and membrane permeability will balance each others and stabilize the in/out metabolites concentration.

EvoEvo Project
FP7-ICT FET Proactive EVLIT program
Project reference: 610427

Deliverable 2.3
Specifications of the population model
Version 1.7

- o Release of DNA (possibly lysed) in environment at death. This could enable for the modelling of transformation (transformation is the process by which a bacterium integrates some exogenous genetic material into its cell).
- o Exchange of plasmids between neighbour cells. This could enable for the modelling of conjugation (conjugation is an active process during which a bacteria injects a part of its genome into another bacteria by cell contact, usually through a sex pilus transferring plasmids).

In all cases, as reminded in deliverable 2.1, specifications can change over time, owing to experimental results, to parametric exploration, interaction with biologists and new insights in the model.

3. Conclusion

The population model is an important component of the EvoEvo modelling effort. By allowing complex interactions between the evolving individuals, it should enable them to *co*-evolve and to develop complex competition or cooperation strategies, or to specialize on the consumption of specific metabolites, possibly released by other organisms (possibly leading to resource cycling). We expect these behaviours to lead to spontaneous environment complexification, due to the presence of different species interacting in different ways and consuming different resources. Ultimately, this complexification could lead to further adaptation to the now complex environment, a process we consider necessary for the population to continuously create new evolutionary challenges, that is to say, to evolve in a open-ended way. The model described in this document additionally allows studying evolution in complex environments (i.e. dynamic non-uniform environment) without direct interaction between the individuals. This will enable iterative study of the evolutionary mechanisms at the level of the population, the environment complexity being predefined or being the result of the organisms' behaviour.

It is important to be aware that the population and environment models designed here are intended to allow for open-endedness. However, there is, to date, no consensus on the definition of open-endedness, much less a methodology to model and simulate it. Consequently, the specifications of the population model may evolve depending on the results of the simulation experiments and on the insights we gain from discussions with other groups or from the literature. In particular, members of the EvoEvo project are involved in a collective discussion on that matter. Although, to date, this discussion does not contradict the choices made here, its progress may lead to refinement of the population model specifications.

4. References

[Ray, 1993] Ray, T. S. (1993). An evolutionary approach to synthetic biology: Zen and the art of creating life. *Artificial Life*, 1(1_2), 179-209.

[Knibbe et al., 2007] Knibbe, C., Coulon, A., Mazet, O., Fayard, J. M., & Beslon, G. (2007). A long-term evolutionary pressure on the amount of noncoding DNA. *Molecular biology and evolution*, 24(10), 2344-2353.

[Crombach & Hogeweg, 2009] Crombach, A., & Hogeweg, P. (2009). Evolution of resource cycling in ecosystems and individuals. *BMC evolutionary biology*, 9(1), 122.

EvoEvo Project
FP7-ICT FET Proactive EVLIT program
Project reference: 610427

Deliverable 2.3
Specifications of the population model
Version 1.7

[Ray, 1991] Ray, T. S. (1991). An approach to the synthesis of life. <u>In:</u> Artificial Life II. Santa Fe Institute Studies in the Sciences of Complexity Vol. X (eds. Langton, C. G., Taylor, C., Farmer, J. D. & Rasmussen, S.) 371–408.

[Mozhayskiy & Tagkopoulos, 2012] Mozhayskiy, V., & Tagkopoulos, I. (2013). Microbial evolution in vivo and in silico: methods and applications. *Integrative Biology*, 5(2), 262-277.

[Hindré et al., 2012] Hindré, T., Knibbe, C., Beslon, G., & Schneider, D. (2012). New insights into bacterial adaptation through in vivo and in silico experimental evolution. *Nature Reviews Microbiology*, 10(5), 352-365.