IN PARTNERSHIP WITH:
Institut national des sciences appliquées de Lyon

Université Claude Bernard (Lyon 1)

Activity Report 2015

Project-Team BEAGLE

Artificial Evolution and Computational Biology

IN COLLABORATION WITH: Laboratoire d'InfoRmatique en Image et Systèmes d'information (LIRIS), Laboratoire de Biométrie et Biologie Evolutive (LBBE), Laboratoire de Recherche en Cardiovasculaire, Métabolisme, Diabétologie et Nutrition

RESEARCH CENTER
Grenoble - Rhône-Alpes

THEME
Computational Biology

Table of contents

Project-Team BEAGLE

Creation of the Team: 2011 June 17, updated into Project-Team: 2013 January 01

Keywords:

Computer Science and Digital Science:

6.1.3. - Discrete Modeling (multi-agent, people centered)
6.1.4. - Multiscale modeling
6.2.7. - High performance computing

Other Research Topics and Application Domains:

1. - Life sciences
1.1.11. - Systems biology
1.1.2. - Molecular biology
1.1.3. - Cellular biology
1.1.8. - Evolutionnary biology
1.1.9. - Bioinformatics

1. Members

Research Scientists

Hugues Berry [Inria, Senior Researcher, HdR]
Eric Tannier [Inria, Researcher, HdR]

Faculty Members

Guillaume Beslon [Team leader, INSA Lyon, Professor, HdR]
Carole Knibbe [Univ. Lyon I, Associate Professor]
Christophe Rigotti [INSA Lyon, Associate Professor, HdR]
Jonathan Rouzaud-Cornabas [INSA Lyon, Associate Professor]
Hedi Soula [INSA Lyon, Associate Professor, HdR]

Engineers

Marie Fernandez [Inria]
Leo Lefebvre [Inria, until Jun 2015]
Vincent Liard [Inria]

PhD Students

Wandrille Duchemin [Univ. Lyon I]
Alexandre Foncelle [Inria]
Alvaro Mateos Gonzalez [ENS Lyon]
Sergio Peignier [Inria]
Ilia Prokin [Inria]
Charles Rocabert [Inria, granted by FP7 EVOEVO project]
Yoram Vadee Le Brun [ENS Lyon]

Post-Doctoral Fellow

Maurizio de Pitta [Inria, IOF Marie Curie Grant]

Visiting Scientist

Priscila Do Nascimento Biller [Universidade Estadual de Campinas, Bresil, until Apr 2015]

Administrative Assistant

Caroline Lothe [Inria]

Others

Raphael Bournhonesque [Inria, Master Student, from Feb 2015 until Jun 2015]
Denis Carnino [Inria, Master Student, from Apr 2015 until Jul 2015]
Barthelemy Caron [Inria, Intern, from Feb 2015 until May 2015]
Audrey Denizot [ENS Lyon, Master Student, from Sep 2015]
Mehdi Ghesh [Inria,Master Student, from May 2015 until Aug 2015]
Theotime Grohens [ENS Paris, Intern, from Jun 2015 until Aug 2015]
Alice Joffard [Inria, Master Student, from Jan until Jun 2015]
André Lamba [Inria, Intern, from Apr 2015 until Jun 2015]
Anthony Rossi [Inria, Intern, from Jun 2015 until Aug 2015]
Jacob Rutten [Inria, Intern, from Nov 2015]
Vladislav Tempez [ENS Rennes, Intern, from May 2015 until Jul 2015]

2. Overall Objectives

2.1. Overall Objectives

The expanded name for the BEAGLE research group is "Artificial Evolution and Computational Biology". Our aim is to position our research at the interface between biology and computer science and to contribute new results in biology by modeling biological systems. In other words we are making artifacts – from the Latin *artis factum* (an entity made by human art rather than by Nature) – and we explore them in order to understand Nature. The team is an Inria Project-Team since January, 2014. It gathers researchers from Inria, INSA, UCBL, who are members of three different labs, the LIRIS [1], the LBBE [2], and CARMEN [3]. It is led by Prof. Guillaume Beslon (INSA-Lyon, LIRIS, Computer Science Dept.).

Our research is based on an interdisciplinary scientific strategy: we are developing computer science formalisms and software for complex system modeling in synergy with multidisciplinary cooperations in the area of life sciences. Using computational approaches we study abstractions of biological systems and processes in order to unravel the organizational principles of cellular systems. More precisely, the scientific activity of the BEAGLE group focuses on two different topics. Both topics are strongly complementary. Indeed, on the short time scales, biological systems are constrained by the physical nature of their substrate but, on long time scales, they are also constrained by their evolutionary history. Thus, studying both time scales and both constraints – including their interactions – gives us a global viewpoint on the roots of biological organization.

Computational Cell Biology We develop models of the spatio-temporal dynamics of cells and their molecular components. More precisely, we study the complex interplay between the reaction and the diffusion processes when the medium is not homogeneous or when the number of molecules is too low to account for a perfect mixing hypothesis. We particularly focus on the consequences on the signaling networks and on the stochasticity of transcription. In this domain, we always try to mix up modeling and "wet" experimental approaches by developing close collaborations with experimental biologists.

Models of Genome Evolution To better understand the cellular structures (genome organization, transcription networks or signaling cascades) we propose to study their historical – evolutionary – origin. Individual-based evolutionary models (*in silico experimental evolution*) allow us to study how evolution leads to some specific structures shaped by the needs of robustness, variability or evolvability, depending on some specific conditions (e.g., large vs. small efficient population sizes, high vs. low mutation rates, stable vs. unstable environments). Models can also be used for predictive purposes

[1] Laboratoire d'Informatique en Image et Systèmes d'Information: UMR 5205 CNRS, INSA-Lyon, Univ. Claude Bernard Lyon 1, Univ. Louis Lumière Lyon 2, École Centrale de Lyon
[2] Laboratoire de Biometrie et Biologie Evolutive: UMR CNRS 5558, Univ. Claude Bernard Lyon 1.
[3] Laboratoire de Recherche en Cardiovasculaire, Métabolisme, Diabétologie et Nutrition: UMR U1060 INSERM, INSA-Lyon, INRA 1235, Univ. Claude Bernard Lyon 1.

on real data: we reconstruct the evolutionary events that have shaped the extant real genomes, including small substitutions as well as large genome reorganizations. By comparing the reconstructed historical events and the laws inferred from artificial experiments, we can explain some patterns of today's organisms and biodiversity.

The scientific objective of the BEAGLE team is to develop a consistent set of concepts and tools – mainly based on computational science – to *in fine* contribute to knowledge discovery in systems biology. Our strategy is to develop strong interactions with life science researchers to become active partners of the biological discovery process. Thus, our aim as a team is not to be a computer science team interacting with biologists, nor to be a team of biologists using computer science tools, but rather to stay in the middle and to become a *trading zone* [54] between biology and computer science. Our very scientific identity is thus fuzzy, melting components from both sciences. Indeed, one of the central claims of the team is that interdisciplinarity involves permanent exchanges between the disciplines. Such exchanges can hardly be maintained between distant teams. That's why the BEAGLE team tries to develop local collaborations with local scientists. That's also why BEAGLE also tries to organize itself as an intrinsically interdisciplinary group, gathering different sensibilities between biology and computer science inside the group. Our ultimate objective is to develop interdisciplinarity at the individual level, all members of the team being able to interact efficiently with specialists from both fields.

3. Research Program

3.1. Introduction

As stated above, the research topics of the BEAGLE Team are centered on the modelisation and simulation of cellular processes. More specifically, we focus on two specific processes that govern cell dynamics and behavior: Evolution and Biophysics. This leads to two main topics: computational cell biology and models for genome evolution.

3.2. Computational Cell Biology

BEAGLE contributes computational models and simulations to the study of cell signaling in prokaryotic and eukaryotic cells, with a special focus on the dynamics of cell signaling both in time and in space. Importantly, our objective here is not so much to produce innovative computer methodologies, but rather to improve our knowledge of the field of cell biology by means of computer methodologies.

This objective is not accessible without a thorough immersion in experimental cell biology. Hence, one specificity of BEAGLE is to be closely associated inside each research project with experimental biology groups. For instance, all the current PhD students implicated in the research projects below have strong interactions with experimenters, most of them conducting experiments themselves in our collaborators' labs. In such a case, the supervision of their PhD is systematically shared between an experimentalist and a theoretician (modeler/computer scientist).

Standard modeling works in cell biochemistry are usually based on mean-field equations, most often referred to as "laws of mass-action". Yet, the derivation of these laws is based on strict assumptions. In particular, the reaction medium must be dilute, perfectly-mixed, three-dimensional and spatially homogeneous and the resulting kinetics are purely deterministic. Many of these assumptions are obviously violated in cells. As already stressed out before, the external membrane or the interior of eukaryotic as well as prokaryotic cells evidence spatial organization at several length scales, so that they must be considered as non-homogeneous media. Moreover, in many case, the small number of molecule copies present in the cell violates the condition for perfect mixing, and more generally, the "law of large numbers" supporting mean-field equations.

When the laws-of-mass-action are invalidated, individual-based models (IBM) appear as the best modeling alternative to evaluate the impact of these specific cellular conditions on the spatial and temporal dynamics of the signaling networks. We develop Individual-Based Models to evaluate the fundamental impact of non-homogeneous space conditions on biochemical diffusion and reaction. More specifically, we focus on the effects of two major sources of non-homogeneity within cells: macromolecular crowding and non-homogeneous diffusion. Macromolecular crowding provides obstacles to the diffusive movement of the signaling molecules, which may in turn have a strong impact on biochemical reactions [42]. In this perspective, we use IBM to renew the interpretation of the experimental literature on this aspect, in particular in the light of the available evidence for anomalous subdiffusion in living cells. Another pertinent source of non-homogeneity is the presence of lipid rafts and/or caveolae in eukaryotic cell membranes that locally alter diffusion. We showed several properties of these diffusion gradients on cells membranes. In addition, combining IBMs and cell biology experiments, we investigate the spatial organization of membrane receptors in plasmic membranes and the impact of these spatial features on the initiation of the signaling networks [46]. More recently, we started to develop IBMs to propose experimentally-verifiable tests able to distinguish between hindered diffusion due to obstacles (macromolecular crowding) and non-homogeneous diffusion (lipid rafts) in experimental data.

The last aspect we tackle concerns the stochasticity of gene expression. Indeed, the stochastic nature of gene expression at the single cell level is now a well established fact [52]. Most modeling works try to explain this stochasticity through the small number of copies of the implicated molecules (transcription factors, in particular). In collaboration with the experimental cell biology group led by Olivier Gandrillon at the Centre de Génétique et de Physiologie Moléculaire et Cellulaire (CGPhyMC, UMR CNRS 5534), Lyon, we study how stochastic gene expression in eukaryotic cells is linked to the physical properties of the cellular medium (e.g., nature of diffusion in the nucleoplasm, promoter accessibility to various molecules, crowding). We have already developed a computer model whose analysis suggests that factors such as chromatin remodeling dynamics have to be accounted for [48]. Other works introduce spatial dimensions in the model, in particular to estimate the role of space in complex (protein+ DNA) formation. Such models should yield useful insights into the sources of stochasticity that are currently not explained by obvious causes (e.g. small copy numbers).

3.3. Models of genome evolution

Classical artificial evolution frameworks lack the basic structure of biological genome (i.e. a double-strand sequence supporting variable size genes separated by variable size intergenic sequences). Yet, if one wants to study how a mutation-selection process is likely (or not) to result in particular biological structures, it is mandatory that the effect of mutation modifies this structure in a realistic way. We have developed an artificial chemistry based on a mathematical formulation of proteins and of the phenotypic traits. In our framework, the digital genome has a structure similar to prokaryotic genomes and a non-trivial genotype-phenotype map. It is a double-stranded genome on which genes are identified using promoter-terminator- like and start-stop-like signal sequences. Each gene is transcribed and translated into an elementary mathematical element (a "protein") and these elements - whatever their number - are combined to compute the phenotype of the organism. The Aevol (Artificial EVOLution) model is based on this framework and is thus able to represent genomes with variable length, gene number and order, and with a variable amount of non-coding sequences (for a complete description of the model, see [59]).

As a consequence, this model can be used to study how evolutionary pressures like the ones for robustness or evolvability can shape genome structure [60], [57], [58], [67]. Indeed, using this model, we have shown that genome compactness is strongly influenced by indirect selective pressures for robustness and evolvability. By genome compactness, we mean several structural features of genome structure, like gene number, amount of non functional DNA, presence or absence of overlapping genes, presence or absence of operons [60], [57], [68]. More precisely, we have shown that the genome evolves towards a compact structure if the rate of spontaneous mutations and rearrangements is high. As far as gene number is concerned, this effect was known as an error-threshold effect [51]. However, the effect we observed on the amount of non functional DNA was unexpected. We have shown that it can only be understood if rearrangements are taken into account: by promoting large duplications or deletions, non functional DNA can be mutagenic for the genes it surrounds.

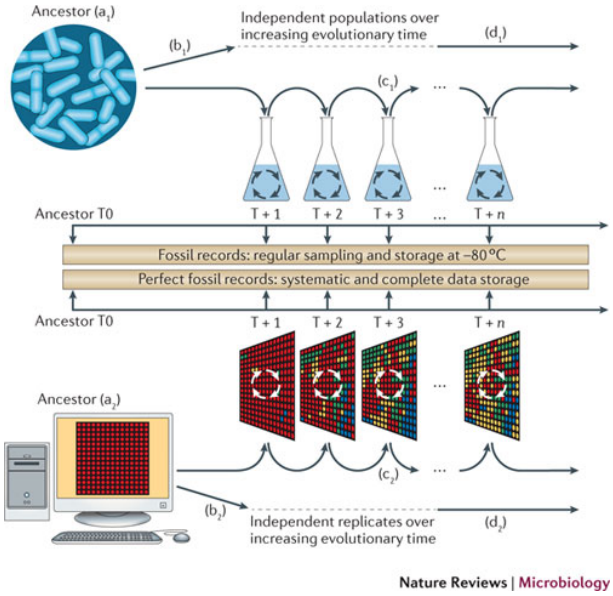

Figure 1. Parallel between experimental evolution and artificial evolution

We have extended this framework to include genetic regulation (R-Aevol variant of the model). We are now able to study how these pressures also shape the structure and size of the genetic network in our virtual organisms [44], [43], [45]. Using R-Aevol we have been able to show that (i) the model qualitatively reproduces known scaling properties in the gene content of prokaryotic genomes and that (ii) these laws are not due to differences in lifestyles but to differences in the spontaneous rates of mutations and rearrangements [43]. Our approach consists in addressing unsolved questions on Darwinian evolution by designing controlled and repeated evolutionary experiments, either to test the various evolutionary scenarios found in the literature or to propose new ones. Our experience is that "thought experiments" are often misleading: because evolution is a complex process involving long-term and indirect effects (like the indirect selection of robustness and evolvability), it is hard to correctly predict the effect of a factor by mere thinking. The type of models we develop are particularly well suited to provide control experiments or test of null hypotheses for specific evolutionary scenarios. We often find that the scenarios commonly found in the literature may not be necessary, after all, to explain the evolutionary origin of a specific biological feature. No selective cost to genome size was needed to explain the evolution of genome compactness [60], and no difference in lifestyles and environment was needed to explain the complexity of the gene regulatory network [43]. When we unravel such phenomena in the individual-based simulations, we try to build "simpler" mathematical models (using for instance population genetics-like frameworks) to determine the minimal set of ingredients required to produce the effect. Both approaches are complementary: the individual-based model is a more natural tool to interact with biologists, while the mathematical models contain fewer parameters and fewer ad-hoc hypotheses about the cellular chemistry.

At this time, simulating the evolution of large genomes during hundreds of thousands of generation with the Aevol software can take several weeks or even months. It is worse with Raevol, where we not only simulate mutations and selection at the evolutionary timescale, but also simulate the lifetime of the individuals, allowing them to respond to environmental signals. Previous efforts to parallelize and distribute Aevol had yielded limited results due to the lack of dedicated staff on these problems. Since September 2014, we have been improving the performance of (R-)Aevol. Thanks to the ADT Aevol, one and a half full time

engineers work on improving Aevol and especially to parallelize it. Moreover, we are working to formalize the numerical computation problems with (R-)Aevol to use state-of-the-art optimization techniques from the HPC community. It ranges from dense and sparse matrix multiplication and their optimizations (such as Tridiagonal matrix algorithm) to using new generation accelerator (Intel Xeon Phi and NVidia GPU). However, our goal is not to become a HPC nor a numerical computation team but to work with well-established teams in these fields, such as through the Joint Laboratory for Extreme-Scale Computing, but also with Inria teams in these fields (e.g. ROMA, Avalon, CORSE, RUNTIME, MESCAL). By doing so, (R-)Aevol simulations will be faster, allowing us to study more parameters in a shorter time. Furthermore, we will also be able to simulate more realistic population sizes, that currently do not fit into the memory of a single computer.

In 2015 we have improved both the quality and the performance of the code. For example, we are currently testing a new representation of the phenotype allowing us to use vector operation. In collaboration with the Avalon team and with the help of a shared internship (Mehdi Ghesh), we have build a benchmark for ordinary differential equation (ODE). This benchmark is based on a representative sample of the ODEs (formalizing the genetic network) found within the R-Aevol model. Thanks to this benchmark, we can compare different ODE solvers and methods. Furthermore, researchers working on ODE solvers and methods could use it to evaluate the quality of their approach. We are now working with Avalon team on an algorithm that will automatically choose at runtime the best fitting solver and method (from a performance and a quality of results point of view). Through this collaboration, we have also extended the execo experimental engine [4] to support Aevol and R-Aevol. By doing so, we have now a complete automatic workflow to conduct large scale campaign experiments with thousands of different parameters of our model and use the resources of distributed platform (Grid'5000 in our case).

Since 2014, we are also working on a second model of genome evolution.This new model, developed by the team within the Evoevo european Project, encompasses not only the gene regulation network (as Raevol does) but also the metabolic level [36]. It allows us to have a real notion of resources and thus to have more complex ecological interactions between the individuals. To speed up computations, the genomic level is simplified compared to aevol, as a chromosome is modelled as a sequence of genes and regulatory elements and not as a sequence of nucleotides. Both models are thus complementary.

Little has been achieved concerning the validation of these models, and the relevance of the observed evolutionary tendencies for living organisms. Some comparisons have been made between Avida and experimental evolution [61], [55], but the comparison with what happened in a long timescale to life on earth is still missing. It is partly because the reconstruction of ancient genomes from the similarities and differences between extant ones is a difficult computational problem which still misses good solutions for every type of mutations, in particular the ones concerning changes in the genome structure.

There exist good phylogenic models of punctual mutations on sequences [53], which enable the reconstruction of small parts of ancestral sequences, individual genes for example [62]. But models of whole genome evolution, taking into account large scale events like duplications, insertions, deletions, lateral transfer, rearrangements are just being developped [70], [49]. Integrative phylogenetic models, considering both nucleotide subsitutions and genome architectures, like Aevol does, are still missing.

Partial models lead to evolutionary hypotheses on the birth and death of genes [50], on the rearrangements due to duplications [41], [69], on the reasons of variation of genome size [56], [63]. Most of these hypotheses are difficult to test due to the difficulty of *in vivo* evolutionary experiments.

To this aim, we develop evolutionary models to reconstruct the history of organisms from the comparison of their genome, at every scale, from nucleotide substitutions to genome organisation rearrangements. These models include large-scale duplications as well as loss of DNA material, and lateral gene transfers from distant species. In particular we have developed models of evolution by rearrangements [64], methods for reconstructing the organization of ancestral genomes [65], [47], [66], or for detecting lateral gene transfer

[4] Matthieu Imbert, Laurent Pouilloux, Jonathan Rouzaud-Cornabas, Adrien Lèbre, Takahiro Hirofuchi "Using the EXECO toolbox to perform automatic and reproducible cloud experiments" 1st International Workshop on UsiNg and building ClOud Testbeds UNICO, collocated with IEEE CloudCom 2013 2013

events [40], [8]. It is complementary with the Aevol development because both the model of artificial evolution and the phylogenetic models we develop emphasize on the architecture of genomes. So we are in a good position to compare artificial and biological data on this point.

We improve the phylogenetic models to reconstruct ancestral genomes, jointly seen as gene contents, orders, organizations, sequences. It requires integrative models of genome evolution, which is desirable not only because they will provide a unifying view on molecular evolution, but also because they will shed light onto the relations between different kinds of mutations, and enable the comparison with artificial experiments from models like Aevol.

Based on this experience, the BEAGLE team contributes individual-based and mathematical models of genome evolution, in silico experiments as well as historical reconstruction on real genomes, to shed light on the evolutionary origin of the complex properties of cells.

4. Application Domains

4.1. Domain

- Genome Evolution
- Computational Systems Biology
- Evolution of Genetic Regulation
- Intracellular Signal Transduction

5. Highlights of the Year

5.1. Highlights of the Year

We organized the first EvoEvo workshop (York, July 2015) as a satellite meeting of the 2015 ECAL conference (http://www.evoevo.eu).

5.1.1. Awards

Best paper award at the ACM Genetic and Evolutionary Computation Conference GECCO'15, in category Evolutionary Machine Learning, for the following paper: .

BEST PAPER AWARD:

[31]
S. PEIGNIER, C. RIGOTTI, G. BESLON. *Subspace Clustering Using Evolvable Genome Structure*, in "Genetic and Evolutionary Computation Conference (GECCO)", Madrid, Spain, July 2015, https://hal.archives-ouvertes.fr/hal-01199136

6. New Software and Platforms

6.1. DeCo

Detection of Co-evolution
KEYWORDS: Bioinformatics - Evolution
SCIENTIFIC DESCRIPTION

The software DeCo computes adjacencies (or any type of relation, like regulation, interaction, functional relationships) between ancestral genes from gene phylogenies reconciled with a species phylogeny according to duplications and losses. It takes as input (1) a species tree (2) a set of extant genes (3) a set of exant adjacencies (relations) between extant genes and (4) gene trees which leaves are the extant genes. It outputs ancestral species, genes, and adjacencies. It also highlights the duplications involving several genes.
FUNCTIONAL DESCRIPTION

DeCo for Detection of Co-evolution, reconstructs neighborhood relationships between genes of ancient genomes, in the presence of gene duplications, transfer and losses.

- Participant: Eric Tannier
- Contact: Eric Tannier
- URL: http://pbil.univ-lyon1.fr/software/DeCo/

6.2. DeCoLT

Detection of Co-evolution with Lateral gene Transfer
KEYWORDS: Bioinformatics - Evolution
SCIENTIFIC DESCRIPTION

The software DeCoLT computes adjacencies (or any type of relation, like regulation, interaction, functional relationships) between ancestral genes from gene phylogenies reconciled with a species phylogeny according to duplications, losses and lateral gene transfer. It takes as input a species tree a set of extant genes a set of exant adjacencies (relations) between extant genes and reconciled gene trees which leaves are the extant genes. It outputs ancestral species, genes, and adjacencies. It also highlights the duplications or transfers involving several genes.
FUNCTIONAL DESCRIPTION

The software DeCoLT computes adjacencies (or any type of relation, like regulation, interaction, functional relationships) between ancestral genes from gene phylogenies reconciled with a species phylogeny according to duplications, losses and lateral gene transfer.

- Participant: Eric Tannier
- Contact: Eric Tannier
- URL: http://pbil.univ-lyon1.fr/software/DeCoLT/

6.3. aevol

Artificial Evolution
FUNCTIONAL DESCRIPTION

Aevol is a digital genetics model: populations of digital organisms are subjected to a process of selection and variation, which creates a Darwinian dynamics. By modifying the characteristics of selection (e.g. population size, type of environment, environmental variations) or variation (e.g. mutation rates, chromosomal rearrangement rates, types of rearrangements, horizontal transfer), one can study experimentally the impact of these parameters on the structure of the evolved organisms. In particular, since Aevol integrates a precise and realistic model of the genome, it allows for the study of structural variations of the genome (e.g. number of genes, synteny, proportion of coding sequences).

The simulation platform comes along with a set of tools for analysing phylogenies and measuring many characteristics of the organisms and populations along evolution.

An extension of the model (R-Aevol), integrates an explicit model of the regulation of gene expression, thus allowing for the study of the evolution of gene regulation networks.

- Participants: Carole Knibbe, Guillaume Beslon, Jonathan Rouzaud-Cornabas, Priscila Do Nascimento Biller, Yoram Vadee Le Brun, David Parsons and Vincent Liard
- Partners: UCBL Lyon 1 - INSERM - Universite Paris-Descartes - Insa de Lyon
- Contact: Carole Knibbe
- URL: http://www.aevol.fr/

6.4. EvoEvo

In silico experimental evolution
KEYWORDS: Bioinformatics - Biology - Evolution
FUNCTIONAL DESCRIPTION

In the context of the EvoEvo european project we are developing an integrated model of microorganisms evolution. This model will extend the current evolutionary models developped in the team (Aevol and R-Aevol) by adding a metabolic level and an ecosystem level. In 2014, a first version has been developed and released that includes the genomic, genetic and metabolic levels.

- Participants: Guillaume Beslon, Charles Rocabert and Carole Knibbe
- Contact: Guillaume Beslon
- URL: http://www.evoevo.eu/

6.5. FluoBacTracker

KEYWORDS: Bioinformatics - Biology - Biomedical imaging
FUNCTIONAL DESCRIPTION FluoBacTracker is an ImageJ () plugin designed to segment and track growing E. Coli cells from microscopy images and movies. FluoBacTracker is a software tool to : i) Select regions of interest in each image (detect the colony), (ii) Denoise and renormalize the images, (iii) Identify each cells in each image (segmentation), (iv) Follow cells through the whole movie (tracking) and (v) Detect divisions and construct cell lineage in the population

- Participants: Magali Vangkeosay, David Parsons and Hugues Berry
- Partner: Universite Descartes
- Contact: Hugues Berry
- URL: http://fluobactracker.inrialpes.fr/

7. New Results

7.1. Subspace Clustering Using Evolvable Genome Structure

We have developped an evolutionary algorithm to tackle the subspace clustering problem. Subspace clustering is recognized as more difficult than standard clustering since it requires to identify not only the clusters but also the various subspaces where the clusters hold. We propose to tackle this problem with a bio-inspired algorithm that includes many bio-like features like variable genome length and organization, functional and non-functional elements, and variation operators including chromosomal rearrangements. These features give the algorithm a large degree of freedom to achieve subspace clustering with satisfying results on a reference benchmark with respect to state of the art methods. One of the main advantages of the approach is that it needs only one subspace clustering ad-hoc parameter: the maximal number of clusters. This is a single and intuitive parameter that sets the maximal level of details of the clustering, while other algorithms require more complicated parameter space exploration. The other parameters of the algorithm are related to the evolution strategy (population size, mutation rate, ...) and for them we use a single setting that turns out to be effective on all the datasets of the benchmark.

This work has been presented at the main conference for genetic & evolutionary computation, GECCO [31], where it received the best paper award and during the EvoEvo Workshop of ECAL 2015 [35].

7.2. Epigenetic inheritance speeds up evolution of artificial organisms

DNA is not the sole medium by which parents transmit information to their offspring. Epigenetic inheritance, in particular, is based on the partial transmission of the cellular state of the parental cell to its descendants. Although the reality of epigenetic inheritance is now firmly established, whether it has an influence on the long term evolutionary process is still subject to debate. To address this question, we used the RAevol extention of the Aevol simulator developed in the team, and defined 4 scenarios with static or dynamic environments and with or without epigenetic inheritance. Simulations in dynamic environments show that protein inheritance indeed increases the rate of evolution on the long term. But they also show that it impedes evolution in its very first stages. This negative effect can be explained by instabilities generated by the interference between the two inheritance mediums. On the opposite, the long term gain can be explained by protein inheritance reducing the constraints on the genetic regulation network.
This work has been published in the article [33].

7.3. In silico evolution improves statistical models of genome dynamics

Using Aevol, we have proved that statistical frameworks published in the last twenty years for inferring evolutionary genome rearrangements are flawed in two ways. First, they mistranslated a null hypothesis on a uniform breakage model, and second, they assumed that genomic breakable regions are known *a priori*. We propose ways to correct these flaws by combining mathematical approaches, simulations, observations and validation on real genomic data. The results will be of interest for an audience from evolutionary biology, computational biology, bioinformatics and mathematics. We successively show that:

- a truly uniform hypothesis on rearrangement breakages leads to a model with an equilibrium intergene size distribution that fits the measured one on diverse genomes,
- estimations based on the flawed uniform breakage model completely fail on simulations with the truly uniform model,
- coherently with previous studies the flawed, and to a lesser extent, the truly uniform model are rejected on amniote genomes if breakable regions are identified with intergenic regions,
- co-estimating the number of breakable regions with the rearrangement distance gives coherent values on amniote genomes.

A paper reporting these results has been submitted by Priscila Biller, Carole Knibbe and Eric Tannier.

7.4. Temperature-induced variation in gene expression burst size in metazoan cells

Gene expression is an inherently stochastic process, owing to its dynamic molecular nature. Protein amount distributions, which can be acquired by cytometry using a reporter gene, can inform about the mechanisms of the underlying microscopic molecular system. By using different clones of chicken erythroid progenitor cells harboring different integration sites of a CMV-driven mCherry protein, we investigated the dynamical behavior of such distributions. We show that, on short term, clone distributions can be quickly regenerated from small population samples with a high accuracy. On longer term, on the contrary, we show variations manifested by correlated fluctuation in the Mean Fluorescence Intensity. In search for a possible cause of this correlation, we demonstrate that in response to small temperature variations cells are able to adjust their gene expression rate: a modest (2 °C) increase in external temperature induces a significant down regulation of mean expression values, with a reverse effect observed when the temperature is decreased. Using a two-state model of gene expression we further demonstrate that temperature acts by modifying the size of transcription bursts, while the burst frequency of the investigated promoter is less systematically affected. For the first time, we report that transcription burst size is a key parameter for gene expression that metazoan cells from homeotherm animals can modify in response to an external thermal stimulus.

This work has been published in the article [11].

7.5. Deciphering the signalling networks of synaptic plasticity

Synaptic plasticity, i.e. adaptive modifications of synaptic strength between two neurons depending on their activity, is a main substrate for learning and memory. Experimentally, synaptic plasticity is commonly assessed using prolonged electrical stimulations. Since learning can arise from few or even a single trial, synaptic strength is expected to adapt rapidly. However, whether synaptic plasticity occurs in response to limited event occurrences remains elusive. To address this question, we started a collaboration with Laurent Venance Lab (experimental neuroscience, College de France, Paris). Combining experimental and modelling approaches, we investigated whether a low number of stimulations can induce plasticity in a major synaptic learning rule, spike-timing-dependent plasticity (STDP). It is known that 100 stimulations induce bidirectional STDP, i.e. spike-timing-dependent potentiation (tLTP) and depression (tLTD) at most central synapses. In rodent striatum, we found that tLTD progressively disappears when the number of stimulations is decreased (below 50 pairings) whereas tLTP displays a biphasic profile: tLTP is observed for 75-100 stimulations, absent for 25-50 stimulations and re-emerges for 5-10 stimulations. This tLTP, induced by very few stimulations (5-10) depends on the endocannabinoid (eCB) system. The eCB system has recently emerged as a pivotal pathway for synaptic plasticity because of its widely characterized ability to depress synaptic transmission on short- and long-term scales. Our result therefore indicate that eCBs also mediate potentiation of the synapse. To understand how eCB signaling may support such bidirectionality, we combined electrophysiology experiments with mathematical modeling. Our model describes the temporal kinetics of the biochemical species involved in a first signaling pathway leading from NMDAR to calmodulin and CaMKII with that of a a second, distinct one that assembles mGluR and cytosolic calcium to eCB production and the resulting activation of CB1R. This demonstrated that STDP outcome is controlled by eCB levels and dynamics: prolonged and moderate levels of eCB lead to eCB-mediated long-term depression (eCB- tLTD) while short and large eCB transients produce eCB-mediated long-term potentiation (eCB-tLTP). Therefore, just like neurotransmitters glutamate or GABA, eCB forms a bidirectional system to encode learning and memory.
For reasons of publication strategy, our first co-publication on the subject presents our major experimental results [16]. A second article, featuring both experimental and modelling results, explains how the underlying signalling network can support the observed bidirectionality and is under submission.

7.6. Anomalous diffusion as an age-structured renewal process

Continuous-time random walks (CTRW) are one of the main mechanisms that are recurrently evoked to explain the emergence of subdiffusion in cells. CTRW were introduced fifty years ago as a generalisation of random walks, where the residence time (the time between two consecutive jumps) is a random variable. If the expectation of the residence time is defined, for instance when it is dirac-distributed or decays exponentially fast, one recovers “normal” Brownian motion. However, when the residence time expectation diverges, the CTRW describes a subdiffusive behavior. The classical approach to CTRW yields a non-Markovian (mean-field) transport equation, which is a serious obstacle when one wants to couple subdiffusion with (bio)chemical reaction. We took an alternative approach to CTRW that maintains the Markovian property of the transport equation at the price of a supplementary independent variable. We associate each random walker with an age a, that is the time elapsed since its last jump and describe the subdiffusive CTRW using an age-structured partial differential equations with age renewal upon each walker jump. In the spatially-homogeneous (zero-dimensional) case, we follow the evolution in time of the age distribution. An approach inspired by relative entropy techniques allows us to obtain quantitative explicit rates for the convergence of the age distribution to a self-similar profile, which corresponds to convergence to a stationary profile for the rescaled variables. An important difficulty arises from the fact that the equation in self-similar variables is not autonomous and we do not have a specific analytical solution. Therefore, in order to quantify the latter convergence, we estimate attraction to a time-dependent “pseudo-equilibrium”, which in turn converges to the stationary profile.
The corresponding article is currently in press [38].

7.7. IGF-I signalling in neural stem cells during neurogenesis and aging

Downregulation of insulin-like growth factor (IGF) pathways prolongs lifespan in various species, including mammals. Still, the cellular mechanisms by which IGF signaling controls the aging trajectory of individual organs are largely unknown. Z. Chaker, in M. Holzenberg Lab (Centre de Recherche Saint-Antoine, Paris), asked whether suppression of IGF-I receptor (IGF-1R) in adult stem cells preserves long-term cell replacement, and whether this may prevent age-related functional decline in a regenerating tissue. Using neurogenesis as a paradigm, we showed that conditional knockout of IGF-1R specifically in adult neural stem cells maintained youthful characteristics of olfactory bulb neurogenesis within an aging brain. This in turn resulted in neuro-anatomical changes that improved olfactory function. To help interpret these results, we developed a mathematical model of stem cell differentiation using ordinary differential equations with time-dependent growth, division and death rates (to account for aging) and optimizing at each time step the amount of IGF-IR to maximize an experimentally-derived tissue efficiency criterion. The model predicts that decreased stimulation of growth in adults is indeed optimal for tissue aging. Thus, inhibiting growth and longevity gene IGF-1R in adult stem cells induced a gain-of-function phenotype during aging, marked by optimized management of cell renewal, and enhanced olfactory sensory function.
This work has been published in the article [14].

7.8. A novel model for leptin resistance

Leptin is a major hormone that regulates food intake and appetite in most mammals. Leptin increase in the blood tends to decrease the food intake and leptin is produced in proportion with fat depot. Leptin is therefore a simple probe that feed backs energy reserve to the brain and maintains a constant weight. It is a central hormone for this balance because KO mice without the leptin gene are quickly extremely obese. Also obese people (and animal) tend to have high concentration of leptin suggesting that after a certain point the brain ignores the leptin signal. We developed a mathematical model that explores this resistance developed by neural cells to leptin. This model predicts leptin resistance if food intake is artificially increased and predict a pathway to obesity by such mechanism. This work has been published by H. Soula (Beagle) in collaboration with F. Crauste (Dracula) with co-supervised PhD student Marine Jacquier[19].

7.9. Without eye contact, birds are Markovian!

Any social birds rely on acoustic messages to organize their daily activity (such as parenting and food foraging). In many occasions, birds are within earshot but not in visual contact and therefore should rely only on acoustic channel for this communication. In collaboration with the University of Saint-Etienne, we developed automatic extraction scripts that can detect birds vocalizations in a protocol of meeting with decreasing distance and with or without visual contact modality. Our worked showed that without visual contact birds are more synchronized and their vocal dynamics cannot be distinguished from a two state Markov chain. This markov property vanishes as soon as visual contact is restablished. This work has been published in the main ethology journal: Animal Behaviour[23].

8. Partnerships and Cooperations

8.1. Regional Initiatives

- Intracell X Evo, projet LABEX ECOFECT. Leaders: Thomas Henry, CIRI, Lyon, and Eric Tannier, Beagle. Other partner : Dominique Schneider, laboratoire Adaptation et pathogénie des Microorganismes, Grenoble. Duration: 3 years The objective of the project is to understand the host-pathogen interactions in the cytosol by an experimental evolution approach. Funding: 120 000 Euros.

8.2. National Initiatives

8.2.1. ANR

- Ancestrome: phylogenetic reconstruction of ancestral "-omes", a five-year project (2012-2017), call "Bioinformatics" of the "Investissements d'avenir". Supervisor: V Daubin (CNRS, LBBE, Lyon) ; with Institut Pasteur, ENS Paris, ISEM (Univ Montpellier 2) Participant: E Tannier.
- Aucomsi (2013-2016) (Models of the vocal tract to study auditory circuits): a 4-year project (2013-2016) funded by a grant from the ANR-NSF-NIH Call for French-US Projects in Computational Neuroscience. With F. Theunissen, UC Berkeley, CA, USA. Supervisor: H. Soula (for France) and F. Theunissen (for US). Participants: H. Soula, M. Fernandez.
- Dopaciumcity: Dopamine modulation of calcium influx underlying synaptic plasticity, a 4-year project (2014-2017) funded by a grant from the ANR-NSF-NIH Call for French-US Projects in Computational Neuroscience. With L. Venance, College de France, CIRB, CNRS/UMR 7241 - INSERM U1050, Paris, France and K Blackwell, Krasnow Institute of Advanced Studies, George Mason University, Fairfax, VA, USA. Supervisor: L Venance (for France) and K.L. Blackwell (for US). Participants: H Berry, I Prokin, A Foncelle

8.3. European Initiatives

8.3.1. FP7 & H2020 Projects

8.3.1.1. EvoEvo

Title: Evolution of Evolution

Programm: FP7

Duration: November 2013 - October 2016

Coordinator: Inria

Partners:

Agencia Estatal Consejo Superior de Investigaciones Cientificas (Spain)

Institut National des Sciences Appliquees de Lyon (France)

Universite Lyon 1 Claude Bernard (France)

Universite Joseph Fourier Grenoble 1 (France)

Universiteit Utrecht (Netherlands)

University of York (United Kingdom)

Inria contact: Guillaume Beslon

Evolution is the major source of complexity on Earth, at the origin of all the species we can observe, interact with or breed. On a smaller scale, evolution is at the heart of the adaptation process for many species, in particular micro-organisms (e.g. bacteria, viruses...). Microbial evolution results in the emergence of the species itself, and it also contributes to the organisms' adaptation to perturbations or environmental changes. These organisms are not only organised by evolution, they are also organised to evolve. The EvoEvo project will develop new evolutionary approaches in information science and will produce algorithms based on the latest understanding of molecular and evolutionary biology. Our ultimate goal is to address open-ended problems, where the specifications are either unknown or too complicated to express, and to produce software able to operate in unpredictable, varying conditions. We will start from experimental observations of micro-organism evolution, and abstract this to reproduce EvoEvo, in biological models, in computational models, and in application software. Our aim is to observe EvoEvo in action, to model EvoEvo, to understand EvoEvo and, ultimately, to implement and exploit EvoEvo in software and computational systems. The EvoEvo project will have impact in ICT, through the development of new technologies. It will also have impact in biology and public health, by providing a better understanding of micro-organism adaptation (such as the emergence of new pathogens or the development of antibiotic resistances).

8.3.1.2. Neuron-Astro-Nets

Title: Neuron-Astro-Nets

Programm: Marie-Curie International Outgoing Fellowship (IOF) grant FP7

Duration: 2013 - October 2017

Coordinator: Inria

Partners:

Inria (France)

Dept Statistics and Neurobiology, University of Chicago (USA)

Inria contact: Hugues Berry

This project aims at developing a new model of synaptic plasticity that takes into account astrocyte signaling, its extension to astrocytes-synapse biochemical interactions in ensembles of synapses enwrapped by the same astrocyte and, eventually, to the firing of a single neuron or networks. The project funds Maurizio De Pitta' s postdoc for 4 years (June 2013- May 2017). M. De Pitta has first spent one year in Beagle, Lyon funded by an EU ERCIM grant (06/2013-05/2014) then two years in N. Brunel's Lab in Chicago (06/2014-05/2016) and one year back in Beagle in Lyon (06/2016-05/2017). The IOF grant funds the last three years.

8.4. International Initiatives

8.4.1. Inria International Partners

8.4.1.1. Informal International Partners

Beagle collaborates with two american laboratories: the Theunissen Lab (UC Berkeley, CA, http://theunissen.berkeley.edu/publications.html) and the Blackwell lab (George Mason Univ., VA, http://krasnow1.gmu.edu/CENlab/index.html). Those labs are the partners of the two ANR-NSF-NIH grants we were awarded (cf "ANR" section above).

8.4.2. Participation In other International Programs

The Beagle team is part of the LIA (Laboratoire International Associé) EvoAct (Evolution in action with living and artificial organisms). EvoAct is a joint laboratory gathering researchers from Dominique Schneider team (UJF, LAPM, UMR CNRS 5163, France), Rich Lenski team (Michigan State University, Beacon center, US) and the Beagle team.

8.5. International Research Visitors

8.5.1. Visits of International Scientists

8.5.1.1. Internships

- Priscila Biller did a one year doctoral internship in Beagle, ending in April 2015
- Jaap Rutten started his internship in the Beagle team in December 2015. Jaap Rutten is a M2 student from the Utrecht University (NL).

8.5.2. Visits to International Teams

8.5.2.1. Research stays abroad

Eric Tannier has spent one month in July 2015 at Simon Fraser University in Vancouver, Canada.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific events organisation

9.1.1.1. General chair, scientific chair

- We organized the first EvoEvo workshop as a satellite meeting of the 2015 ECAL conference (http://www.evoevo.eu), July, York, UK (G. Beslon, C. Knibbe, co-organizers).
- Workshop "Molecule Trajectories in Cellular Spaces: promoting interactions between theoreticians and experimentalists" (http://traece.inria.fr), November, Lyon (H. Berry, organizer)
- Annual CNRS-INRA thematic school "EIEFB: Ecole interdisciplinaire d'echanges et de formation en biologie" (http://ecoleporquerolles.inria.fr), June, Villers-sur-Mer (H. Berry, co-organizer)

9.1.1.2. Member of the organizing committees

- Bis-annual CNRS thematic school "CompSysBio: Advanced Lecture Course on Computational Systems Biology" (http://compsysbio.inria.fr), Autrans (H. Berry, G. Beslon)
- "EvoLyon 2015: Conference Lyonnaise sur l'evolution" (http://evolyon.universite-lyon.fr), Lyon (G. Beslon)
- International Conference on Systems Biology, "LyonSysBio 2015" (http://lyonsysbio2015.sciencesconf.org/?lang=en), Lyon (G. Beslon)

9.1.2. Scientific events selection

9.1.2.1. Member of the conference program committees

- Conférence en Parallélisme, Architecture et Système "COMPASS" (http://compas15.lifl.fr/) Lille (J. Rouzaud-Cornabas)
- 13th European Conference on Artificial Life "ECAL 2015" (https://www.cs.york.ac.uk/nature/ecal2015/), York, UK (G. Beslon, C. Knibbe)
- International Conference on Data Mining (ICDM) (http://icdm2015.stonybrook.edu), Atlantic City, NJ, USA (C. Rigotti)
- RECOMB-Comparative Genomics, 2015 (https://applbio.biologie.uni-frankfurt.de/recombcg2015/), Frankfurt, Germany (E. Tannier)

9.1.2.2. Reviewer

- Euro-Par 2015, Vienna, Austria (J. Rouzaud-Cornabas)
- RECOMB-Comparative Genomics 2015 (E. Tannier)
- RECOMB-2016 (E. Tannier)

9.1.3. Journal

9.1.3.1. Member of the editorial boards

- Journal of Complex Systems, AIMS Biophysics (H. Berry)

9.1.3.2. Reviewer - Reviewing activities

In 2015, Beagle members have reviewed numerous papers for international journals, including PLoS Comput Biol, Frontiers Synaptic Neuroscience, Frontiers Computational Neuroscience, J Comput Neurosci, J Theor Biol, New J Physics, Systematic Biology, BMC Bioinformatics, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing....

9.1.4. Invited talks

- H. Berry gave invited talks at the Quantitative BioImaging 2015 conference (Paris), the NYU Adu-Dhabi Workshop on Computational Neuroscience (Abu-Dhabi) and the Neuron-Glia Interactions Workshop of the European Institute for Theoretical Neuroscience (Paris) as well as invited seminars in the following labs / groups: Center for Mathematical Medecine and Biology (Nottingham, UK), Phlam (Lille), Laboratoire de bioenergetique fondamentale et appliquee (Grenoble) and Centre de Recherche en Neurobiologie et Neurophysiologie (Marseille).
- G. Beslon has been invited to give seminars at the Basel Biozentrum (Switzerland), at the UTH Zurich systems biology doctoral program (Switzerland), at the Evolutionary Algorithms conference (Lyon), at the BioVision days (Lyon) and at the Laboratoire de Biologie Quantitative et Computationnelle (UPMC, Paris).
- H. Soula gave an invited talk to the Workshop “Molecule Trajectories in Cellular Spaces: promoting interactions between theoreticians and experimentalists” in Lyon and an invited seminar at the Laboratoire de Recherche en Informatique (LRI), Orsay.
- E. Tannier was invited to give a series of 4 lectures at Simon Fraser University, Vancouver, Canada and gave invited talks to the "Phylogenetic Networks" conference (Singapore). He was also invited to give seminars at the Biosticker seminar (Nantes), to the G-Scop lab (Grenoble) and during the interdisciplinary meeting "evolution of genomes and languages" (Lyon).

9.1.5. Scientific expertise

- ANR call “Numerique et Societe” (C. Rigotti)
- Evaluation committee for the calls for funding “Systems Biology and cancer” of the “ITMO Cancer” (H. Berry)
- Science Steering Committee of the Rhone-Alpes Complex Systems Institute (IXXI) (http://www.ixxi.fr) (H. Berry)
- Scientific board of CNRS GdR MIV (Microscopie et Imagerie du Vivant, GdR 2588, http://gdr-miv.fr) (H. Berry)

9.1.6. Research administration

- Comite National de la Recherche Scientifique, CNRS, Section 6 (G. Beslon, member)
- Comite National de la Recherche Scientifique, CNRS, Section 51 (G. Beslon, member)
- Inria Administration council (Conseil d’Administration, E. Tannier, member)
- Inria Scientific board (Conseil scientifique, H. Berry, member)
- Inria “Parity-Equality” committee (H. Berry, member)
- Inria “Symposium committee” (E. Tannier, scientific director)
- Inria Grenoble-Rhone Alpes hiring committee for “young researchers" (CR2) (H. Berry, head of the committee)
- Inria Grenoble-Rhone Alpes Comite de Developpement Technologique (CDT) (C. Knibbe, member)
- Selection committee of section 64 at INSA de Lyon (H. Soula)
- Inria hiring committee for “senior researchers" (DR2) (H. Berry, member)
- Conseil de Laboratoire LIRIS lab (UMR 5205 CNRS) (G. Beslon, C. Knibbe, members)

9.2. Teaching - Supervision - Juries

9.2.1. Supervision

In addition to the PhD students of the group, listed above, Beagle members have been co-supervising PhD students from other groups. In 2015 E. Tannier has been co-supervising the PhD of Y. Anselmetti (Isem, Montpellier, co-supervised with Severine Berard) and M. Semeria (LBBE, Lyon, co-supervised with Laurent Gueguen). E. Tannier has also supervised P. Do Nascimento Biller’s one-year doctoral internship (Universidade Estadual de Campinas, Brazil) in Beagle.

9.2.2. Juries

9.2.2.1. PhD Juries

- A. Garnier, Univ. P & M Curie, Paris, December 2015 (H. Berry, rapporteur)
- A. Grignard, Univ. P & M Curie, Paris, Octobre 2015 (G. Beslon, rapporteur)
- N. Subramaniyam, Tampere University of Technology, Finland, December 2015 (H. Berry, rapporteur)
- L. Viraphong-Caudwell, Universite Grenoble Alpes, Octobre 2015 (G. Beslon, Rapporteur)
- Z. Yekuieii, Tel Aviv University, Israel, June 2015 (H. Soula, rapporteur)

9.2.2.2. HDR Juries

P. Redou, University of Brest, France, June 2015(H. Soula, rapporteur)

9.3. Popularization

- Eric Tannier gave a series of three lectures to the “Université populaire de Lyon”, entitled “Biocritique”

10. Bibliography

Major publications by the team in recent years

[1] A.-S. COQUEL, J.-P. JACOB, M. PRIMET, A. DEMAREZ, M. DIMICCOLI, T. JULOU, L. MOISAN, A. B. LINDNER, H. BERRY. *Localization of protein aggregation in Escherichia coli is governed by diffusion and nucleoid macromolecular crowding effect*, in "PLoS Computational Biology", 2013, vol. 9, n^{o} 4 [*DOI :* 10.1371/JOURNAL.PCBI.100303], http://hal.inria.fr/hal-00798053

[2] Y. CUI, V. PAILLE, H. XU, S. GENET, B. DELORD, E. FINO, H. BERRY, L. VENANCE. *Endocannabinoids mediate bidirectional striatal spike-timing dependent plasticity*, in "Journal of Physiology", 2015, vol. 593, n^{o} 13, pp. 2833-2849 [*DOI :* 10.1113/JP270324], https://hal.inria.fr/hal-01141205

[3] W. DUCHEMIN, V. DAUBIN, E. TANNIER. *Reconstruction of an ancestral Yersinia pestis genome and comparison with an ancient sequence*, in "BMC Genomics", 2015, vol. 16, n^{o} Suppl 10, S9 p. [*DOI :* 10.1007/978-3-319-19048-8_16], https://hal.inria.fr/hal-01179197

[4] S. FISCHER, S. BERNARD, G. BESLON, C. KNIBBE. *A model for genome size evolution*, in "Bulletin of Mathematical Biology", September 2014, vol. 76, n^{o} 9, pp. 2249-2291 [*DOI :* 10.1007/S11538-014-9997-8], https://hal.archives-ouvertes.fr/hal-01090964

[5] T. HINDRÉ, C. KNIBBE, G. BESLON, D. SCHNEIDER. *New insights into bacterial adaptation through in vivo and in silico experimental evolution*, in "Nature Reviews Microbiology", 2012, vol. 10, pp. 352-365, http://hal.inria.fr/hal-00696231

[6] P.-N. MOUGEL, C. RIGOTTI, M. PLANTEVIT, O. GANDRILLON. *Finding maximal homogeneous clique sets*, in "Knowledge and Information Systems", March 2013, vol. 35, n^{o} 1, pp. 1-30 [*DOI :* 10.1007/S10115-013-0625-Y], http://hal.inria.fr/hal-00827164

[7] H. SOULA, B. CARÉ, G. BESLON, H. BERRY. *Anomalous versus slowed-down Brownian diffusion in the ligand-binding equilibrium*, in "Biophysical Journal", 2013, vol. 105, n^{o} 9, pp. 2064-2073 [*DOI :* 10.1016/J.BPJ.2013.07.023], http://hal.inria.fr/hal-00720515

[8] G. J. SZÖLLOSI, B. BOUSSAU, S. S. ABBY, E. TANNIER, V. DAUBIN. *Phylogenetic modeling of lateral gene transfer reconstructs the pattern and relative timing of speciations*, in "Proceedings- National Academy of Sciences Usa", October 2012, vol. 109, n° 43, pp. 17513-17518 [*DOI :* 10.1073/PNAS.1202997109], http://hal.inria.fr/hal-00740292

[9] J. VIÑUELAS, G. KANEKO, A. COULON, E. VALLIN, V. MORIN, C. MEJIA-POUS, J.-J. KUPIEC, G. BESLON, O. GANDRILLON. *Quantifying the contribution of chromatin dynamics to stochastic gene expression reveals long, locus-dependent periods between transcriptional bursts*, in "BMC Biology", February 2013, vol. 11, n° 1, 15 p. [*DOI :* 10.1186/1741-7007-11-15], http://hal.inria.fr/inserm-00817963

Publications of the year

Articles in International Peer-Reviewed Journals

[10] Y. ANSELMETTI, V. BERRY, C. CHAUVE, A. CHATEAU, E. TANNIER, S. BÉRARD. *Ancestral gene synteny reconstruction improves extant species scaffolding*, in "BMC Genomics", 2015, vol. 16, n° Suppl 10, S11 [*DOI :* 10.1101/023085], https://hal.archives-ouvertes.fr/hal-01180303

[11] O. ARNAUD, S. MEYER, E. VALLIN, G. BESLON, O. GANDRILLON. *Temperature-induced variation in gene expression burst size in metazoan cells*, in "BMC Molecular Biology", November 2015, vol. 16, n° 20 [*DOI :* 10.1186/S12867-015-0048-2], https://hal.inria.fr/hal-01248384

[12] P. BILLER, L. GUÉGUEN, E. TANNIER. *Moments of genome evolution by Double Cut-and-Join*, in "BMC Bioinformatics", 2015, vol. 16, n° Suppl 14, S7 [*DOI :* 10.1186/1471-2105-16-S14-S7], https://hal.inria.fr/hal-01179597

[13] P. CELLIER, T. CHARNOIS, M. PLANTEVIT, C. RIGOTTI, B. CRÉMILLEUX, O. GANDRILLON, J. KLEMA, J.-L. MANGUIN. *Sequential pattern mining for discovering gene interactions and their contextual information from biomedical texts*, in "Journal of Biomedical Semantics", 2015, vol. 6, 27 p. [*DOI :* 10.1186/S13326-015-0023-3], https://hal.archives-ouvertes.fr/hal-01192959

[14] Z. CHAKER, S. AÏD, H. BERRY, M. HOLZENBERGER. *Suppression of IGF-I signals in neural stem cells enhances neurogenesis and olfactory function during aging*, in "Aging Cell", July 2015, 10 p. [*DOI :* 10.1111/ACEL.12365], https://hal.inria.fr/hal-01163564

[15] F. CRAUSTE, E. TERRY, I. L. MERCIER, J. MAFILLE, S. DJEBALI, T. ANDRIEU, B. MERCIER, G. KANEKO, C. ARPIN, J. MARVEL, O. GANDRILLON. *Predicting pathogen-specific CD8 T cell immune responses from a modeling approach*, in "Journal of Theoretical Biology", June 2015, vol. 374, pp. 66-82 [*DOI :* 10.1016/J.JTBI.2015.03.033], https://hal.archives-ouvertes.fr/hal-01242319

[16] Y. CUI, V. PAILLE, H. XU, S. GENET, B. DELORD, E. FINO, H. BERRY, L. VENANCE. *Endocannabinoids mediate bidirectional striatal spike-timing dependent plasticity*, in "Journal of Physiology", 2015, vol. 593, n° 13, pp. 2833-2849 [*DOI :* 10.1113/JP270324], https://hal.inria.fr/hal-01141205

[17] W. DUCHEMIN, V. DAUBIN, E. TANNIER. *Reconstruction of an ancestral Yersinia pestis genome and comparison with an ancient sequence*, in "BMC Genomics", 2015, vol. 16, n° Suppl 10, S9 [*DOI :* 10.1007/978-3-319-19048-8_16], https://hal.inria.fr/hal-01179197

[18] A. GELOEN, H. SOULA, L. HADJI, E. BERGER. *Hysteresis of White Adipose Tissue*, in "Open Journal of Endocrine and Metabolic Diseases", 2015, vol. 5, n° 10, 138 p. , https://hal.archives-ouvertes.fr/hal-01253148

[19] M. JACQUIER, H. A. SOULA, F. CRAUSTE. *A mathematical model of leptin resistance*, in "Mathematical Biosciences", September 2015, vol. 267 [*DOI :* 10.1016/J.MBS.2015.06.008], https://hal.inria.fr/hal-01233483

[20] C. LANGLOIS, T. DOUILLARD, H. YUAN, N. P. BLANCHARD, A. DESCAMPS-MANDINE, B. VAN DE MOORTÈLE, C. RIGOTTI, T. EPICIER. *Crystal orientation mapping via ion channeling: An alternative to EBSD*, in "Ultramicroscopy", October 2015, vol. 157, pp. 65-72 [*DOI :* 10.1016/J.ULTRAMIC.2015.05.023], https://hal.archives-ouvertes.fr/hal-01226846

[21] F. MURAT, R. R. ZHANG, S. GUIZARD, H. GAVRANOVIC, R. FLORES, D. STEINBACH, H. QUESNEVILLE, E. TANNIER, J. SALSE. *Karyotype and Gene Order Evolution from ReconstructedExtinct Ancestors Highlight Contrasts in Genome Plasticity ofModern Rosid Crops*, in "Genome Biology and Evolution", 2015, vol. 7, n° 3, pp. 735-749 [*DOI :* 10.1093/GBE/EVV014], https://hal.archives-ouvertes.fr/hal-01133906

[22] E. C. PEREZ, J. E. ELIE, I. C. BOUCAUD, T. CROUCHET, C. O. SOULAGE, H. A. SOULA, F. E. THEUNISSEN, C. VIGNAL. *Physiological resonance between mates through calls as possible evidence of empathic processes in songbirds*, in "Hormones and Behavior", 2015, vol. 75, pp. 130–141, https://hal.archives-ouvertes.fr/hal-01253147

[23] E. PEREZ, M. FERNANDEZ, S. GRIFFITH, C. VIGNAL, H. A. SOULA. *Impact of visual contact on vocal interaction dynamics of pair-bonded birds*, in "Animal Behaviour", 2015, vol. 107, pp. 125–137, https://hal.archives-ouvertes.fr/hal-01253146

[24] M. SEMERIA, E. TANNIER, L. GUÉGUEN. *Probabilistic modeling of the evolution of gene synteny within reconciled phylogenies*, in "BMC Bioinformatics", 2015, vol. 16, n° Suppl 14, S5 [*DOI :* 10.1186/1471-2105-16-S14-S5], https://hal.inria.fr/hal-01179596

[25] H. SOULA, A. GÉLOËN, C. SOULAGE. *Model of adipose tissue cellularity dynamics during food restriction*, in "Journal of Theoretical Biology", January 2015, vol. 364, 10 p. [*DOI :* 10.1016/J.JTBI.2014.08.046], https://hal.inria.fr/hal-01092535

[26] G. SZOLLOSI, A. A. DAVIN, E. TANNIER, V. DAUBIN, B. BOUSSAU. *Genome-scale phylogenetic analysis finds extensive gene transfer among fungi*, in "Philosophical transactions of the Royal Society of London. Series B, Biological sciences", 2015 [*DOI :* 10.1098/RSTB.2014.0335], https://hal.archives-ouvertes.fr/hal-01218565

[27] G. SZÖLLŐSI, E. TANNIER, V. DAUBIN, B. BOUSSAU. *The inference of gene trees with species trees*, in "Systematic biology", 2015, vol. 64, n° 1, pp. e42-e62 [*DOI :* 10.1093/SYSBIO/SYU048], https://hal.archives-ouvertes.fr/hal-00915301

International Conferences with Proceedings

[28] A. BOUSQUET, J. BRIFFAUT, E. CARON, E. MARÍA DOMINGUEZ, J. FRANCO, A. LEFRAY, O. LÓPEZ, S. ROS, J. ROUZAUD-CORNABAS, C. TOINARD, M. URIARTE. *Enforcing Security and Assurance Properties in Cloud Environment*, in "8th IEEE/ACM International Conference on Utility and Cloud Computing (UCC 2015)", Limassol, Cyprus, 8th IEEE/ACM International Conference on Utility and Cloud Computing, University of Cyprus, December 2015, https://hal.inria.fr/hal-01240557

[29] A. LEFRAY, E. CARON, J. ROUZAUD-CORNABAS, C. TOINARD. *Microarchitecture-Aware Virtual Machine Placement under Information Leakage Constraints*, in "8th IEEE International Conference on Cloud Comput-

ing (IEEE Cloud 2015)", New-York, United States, June 2015, n^{O} Print ISBN: 978-1-4673-7286-2, pp. 588 - 595, http://www.thecloudcomputing.org/2015/ [*DOI :* 10.1109/CLOUD.2015.84], https://hal.inria.fr/hal-01240573

[30] N. MÉGER, C. RIGOTTI, C. POTHIER. *Swap Randomization of Bases of Sequences for Mining Satellite Image Time Series*, in "European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)", Porto, Portugal, Lecture Notes in Computer Science volume 9285, September 2015, https://hal.archives-ouvertes.fr/hal-01200674

[31] *Best Paper*
S. PEIGNIER, C. RIGOTTI, G. BESLON. *Subspace Clustering Using Evolvable Genome Structure*, in "Genetic and Evolutionary Computation Conference (GECCO)", Madrid, Spain, July 2015, https://hal.archives-ouvertes.fr/hal-01199136.

[32] Y. PERICAULT, C. POTHIER, N. MEGER, C. RIGOTTI, F. VERNIER, H.-T. PHAM, E. TROUVE. *A swap randomization approach for mining motion field time series over the Argentiere glacier*, in "2015 8th International Workshop on the Analysis of Multitemporal Remote Sensing Images (Multi-Temp)", Annecy, France, July 2015 [*DOI :* 10.1109/MULTI-TEMP.2015.7245757], https://hal.archives-ouvertes.fr/hal-01237206

[33] Y. VADÉE-LE-BRUN, J. ROUZAUD-CORNABAS, G. BESLON. *Epigenetic inheritance speeds up evolution of artificial organisms*, in "European Conference on Artificial Life", York, United Kingdom, Proceedings of the European Conference on Artificial Life, July 2015 [*DOI :* 10.7551/978-0-262-33027-5-CH078], https://hal.inria.fr/hal-01248383

Conferences without Proceedings

[34] J. GIPPET, C. ROCABERT, S. FENET, A. DUMET, B. KAUFMANN. *Modeling and evaluating human-mediated dispersal mechanisms at landscape scale: a study of road network and invation parameters for Lasius neglectus ants invasive species*, in "World Conference on Natural Resource Modeling", Bordeaux, France, June 2015, https://hal.archives-ouvertes.fr/hal-01242828

[35] S. PEIGNIER, C. RIGOTTI, G. BESLON. *Subspace Clustering for all Seasons*, in "EvoEvo Workshop (satellite workshop of ECAL 2015)", york, United Kingdom, July 2015, https://hal.inria.fr/hal-01252793

[36] C. ROCABERT, C. KNIBBE, G. BESLON. *Towards a Integrated Evolutionary Model to Study Evolution of Evolution*, in "EvoEvo Workshop (Satellite workshop of ECAL 2015)", York, United Kingdom, July 2015, https://hal.inria.fr/hal-01252796

Scientific Books (or Scientific Book chapters)

[37] J.-G. DUMAS, J.-L. ROCH, E. TANNIER, S. VARRETTE. *Foundations of Coding: Compression, Encryption, Error-Correction*, Wiley, 2015, 376 p. , https://hal.archives-ouvertes.fr/hal-00765802

Other Publications

[38] H. BERRY, T. LEPOUTRE, Á. MATEOS GONZÁLEZ. *Quantitative convergence towards a self similar profile in an age-structured renewal equation for subdiffusion*, March 2015, working paper or preprint, https://hal.inria.fr/hal-01136667

[39] E. NOUTAHI, M. SEMERIA, M. LAFOND, J. SEGUIN, B. BOUSSAU, L. GUÉGUEN, N. EL-MABROUK, E. TANNIER. *Efficient gene tree correction guided by species and synteny evolution*, August 2015, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01162963

References in notes

[40] S. S. ABBY, E. TANNIER, M. GOUY, V. DAUBIN. *Detecting lateral gene transfers by statistical reconciliation of phylogenetic forests*, in "BMC Bioinformatics", 2010, vol. 11, 13 p. , http://dx.doi.org/10.1186/1471-2105-11-324

[41] J. A. BAILEY, R. BAERTSCH, W. J. KENT, D. HAUSSLER, E. E. EICHLER. *Hotspots of mammalian chromosomal evolution*, in "Genome Biol", 2004, vol. 5, n° 4, 7 p. , http://dx.doi.org/10.1186/gb-2004-5-4-r23

[42] H. BERRY. *Monte Carlo simulations of enzyme reactions in two dimensions: fractal kinetics and spatial segregation*, in "Biophys J", 2002, vol. 83, n° 4, pp. 1891–1901

[43] G. BESLON, D. P. PARSONS, Y. SANCHEZ-DEHESA, J.-M. PEÑA, C. KNIBBE. *Scaling Laws in Bacterial Genomes: A Side-Effect of Selection of Mutational Robustness*, in "BioSystems", 2010, vol. 102, n° 1, pp. 32-40

[44] G. BESLON, Y. SANCHEZ-DEHESA, D. P. PARSONS, J.-M. PEÑA, C. KNIBBE. *Scaling Laws in Digital Organisms*, in "Proceedings of Information Processing in Cells and Tissues (IPCAT'09)", 2009, pp. 111-114

[45] G. BESLON, Y. SANCHEZ-DEHESA, D. P. PARSONS, C. RIGOTTI, J.-M. PEÑA. *From Digital Genetics to Knowledge Discovery: Perspectives in Genetic Network Understanding*, in "Intelligent Data Analysis journal (IDAj)", 2010, vol. 14, n° 2, pp. 173-191

[46] B. CARÉ, H. A. SOULA. *Impact of receptor clustering on ligand binding*, in "BMC Systems Biology", March 2011, vol. 5, n° 1, 48 p. , PMID: 21453460 [*DOI :* 10.1186/1752-0509-5-48], http://www.ncbi.nlm.nih.gov/pubmed/21453460

[47] C. CHAUVE, H. GAVRANOVIC, A. OUANGRAOUA, E. TANNIER. *Yeast ancestral genome reconstructions: the possibilities of computational methods II*, in "J Comput Biol", Sep 2010, vol. 17, n° 9, pp. 1097–1112, http://dx.doi.org/10.1089/cmb.2010.0092

[48] A. COULON, O. GANDRILLON, G. BESLON. *On the spontaneous stochastic dynamics of a single gene: complexity of the molecular interplay at the promoter*, in "BMC Systems Biology", 2010, vol. 4, n° 1, 2 p.

[49] A. E. DARLING, I. MIKLÓS, M. A. RAGAN. *Dynamics of genome rearrangement in bacterial populations*, in "PLoS Genet", 2008, vol. 4, n° 7, e1000128, http://dx.doi.org/10.1371/journal.pgen.1000128

[50] L. A. DAVID, E. J. ALM. *Rapid evolutionary innovation during an Archaean genetic expansion*, in "Nature", Jan 2011, vol. 469, n° 7328, pp. 93–96 [*DOI :* 10.1038/NATURE09649]

[51] M. EIGEN. *Selforganization of matter and the evolution of biological macromolecules*, in "Naturwissenschaften", 1971, vol. 58, n° 10, pp. 465-523

[52] M. ELOWITZ, A. LEVINE, E. SIGGIA, P. SWAIN. *Stochastic gene expression in a single cell*, in "Science", 2002, vol. 297, n^{o} 5584, pp. 1183–1186

[53] J. FELSENSTEIN. *Inferring phylogenies*, Sinauer Associates, 2004

[54] P. GALISON. *Image and Logic: A Material Culture of Microphysics*, University Of Chicago Press, 1997

[55] T. HINDRÉ, C. KNIBBE, G. BESLON, D. SCHNEIDER. *New insights into bacterial adaptation through in vivo and in silico experimental evolution*, in "Nature Reviews Microbiology", 2012, vol. 10, pp. 352-365, http://hal.inria.fr/hal-00696231

[56] INTERNATIONAL APHID GENOMICS CONSORTIUM. *Genome sequence of the pea aphid Acyrthosiphon pisum*, in "PLoS Biol", Feb 2010, vol. 8, n^{o} 2, e1000313, http://dx.doi.org/10.1371/journal.pbio.1000313

[57] C. KNIBBE, A. COULON, J.-M. FAYARD, G. BESLON. *A long term evolutionary pressure on the amount of noncoding DNA*, in "Molecular Biology and Evolution", 2007, vol. 24, n^{o} 10, pp. 2344-2353

[58] C. KNIBBE, J.-M. FAYARD, G. BESLON. *The topology of the protein network influences the dynamics of gene order : From systems biology to a systemic understanding of evolution*, in "Artificial Life", 2008, vol. 14, n^{o} 1, pp. 149-156

[59] C. KNIBBE. *Structuration de génomes par sélection indirecte de la variabilité mutationnelle, une approche par modélisation et simulation*, PhD Thesis, Institut National des Sciences Appliquées de Lyon, 2006, 174 p.

[60] C. KNIBBE, O. MAZET, F. CHAUDIER, J.-M. FAYARD, G. BESLON. *Evolutionary coupling between the deleteriousness of gene mutations and the amount of non-coding sequences*, in "Journal of Theoretical Biology", 2007, vol. 244, n^{o} 4, pp. 621–630

[61] R. E. LENSKI, C. OFRIA, R. T. PENNOCK, C. ADAMI. *The evolutionary origin of complex features*, in "Nature", 2003, vol. 423, pp. 139–144

[62] D. A. LIBERLES. *Ancestral Sequence Reconstruction*, Oxford University Press, 2007

[63] G. A. B. MARAIS, A. FORREST, E. KAMAU, J. KÄFER, V. DAUBIN, D. CHARLESWORTH. *Multiple nuclear gene phylogenetic analysis of the evolution of dioecy and sex chromosomes in the genus Silene*, in "PLoS One", 2011, vol. 6, n^{o} 8, e21915, http://dx.doi.org/10.1371/journal.pone.0021915

[64] I. MIKLÓS, E. TANNIER. *Bayesian sampling of genomic rearrangement scenarios via double cut and join*, in "Bioinformatics", Dec 2010, vol. 26, n^{o} 24, pp. 3012–3019, http://dx.doi.org/10.1093/bioinformatics/btq574

[65] F. MURAT, J.-H. XU, E. TANNIER, M. ABROUK, N. GUILHOT, C. PONT, J. MESSING, J. SALSE. *Ancestral grass karyotype reconstruction unravels new mechanisms of genome shuffling as a source of plant evolution*, in "Genome Res", Nov 2010, vol. 20, n^{o} 11, pp. 1545–1557 [*DOI :* 10.1101/GR.109744.110]

[66] A. OUANGRAOUA, E. TANNIER, C. CHAUVE. *Reconstructing the architecture of the ancestral amniote genome*, in "Bioinformatics", Oct 2011, vol. 27, n^{o} 19, pp. 2664–2671, http://dx.doi.org/10.1093/bioinformatics/btr461

[67] D. P. PARSONS, C. KNIBBE, G. BESLON. *Importance of the rearrangement rates on the organization of transcription*, in "Proceedings of Artificial Life 12", MIT Press, 2010, pp. 479-486

[68] D. P. PARSONS, C. KNIBBE, G. BESLON. *Homologous and nonhomologous rearrangements: Interactions and effects on evolvability*, in "European Conference on Artificial Life (ECAL)", MIT Press, 2011, pp. 622–629

[69] M. SÉMON, K. H. WOLFE. *Consequences of genome duplication*, in "Curr Opin Genet Dev", Dec 2007, vol. 17, n^{o} 6, pp. 505–512, http://dx.doi.org/10.1016/j.gde.2007.09.007

[70] A. TOFIGH, M. HALLETT, J. LAGERGREN. *Simultaneous identification of duplications and lateral gene transfers*, in "IEEE/ACM Trans Comput Biol Bioinform", 2011, vol. 8, n^{o} 2, pp. 517–535, http://dx.doi.org/10.1109/TCBB.2010.14