

## INTEGRATED PROJECT “DNA & GENOMES”

**KEYWORDS:** HIGH-THROUGHPUT SEQUENCING, FUNCTIONAL GENOMICS, STRUCTURAL GENOMICS, POPULATION GENETICS, PHYLOGENOMICS, EXOME, TRANSCRIPTOME, PROTEOME, EPIGENETICS, PHARMACOGENOMICS, PAN-GENOME, METAGENOME, ENVIRONMENTAL GENOMICS, MATHEMATICAL MODELING, STATISTICAL INFERENCE, BIOPHYSICS, BIOINFORMATICS.

**COORDINATOR:** F. PARDI (LIRMM)

**MAJOR RESEARCH UNITS INVOLVED:** LIRMM, IMAG, L2C

**PHD FUNDED 2015-2017:** 3

**MAN-MONTHS POST-DOC FUNDED 2015-2017:** 66

**EXAMPLES OF CROSS COOPERATION:** BANYULS OCEANOLOGICAL OBSERVATORY; UMR CBS: CENTRE DE BIOCHIMIE STRUCTURAL; UMR LABORATOIRE DE MICROBIOLOGIE ET GÉNÉTIQUE MOLÉCULAIRE (TOULOUSE); UMR ISEM: INSTITUT DES SCIENCES DE L'ÉVOLUTION DE MONTPELLIER, RESEARCH ACTIONS WITH LABEX EPINGENMED (MONTPELLIER) AND LABEX CEMEB (MONTPELLIER); INSTITUT PASTEUR (PARIS); IBC: INSTITUT DE BIOLOGIE COMPUTATIONNELLE (MONTPELLIER)

**EXAMPLES OF INTERNATIONAL COOPERATION:** JOINT GENOME INSTITUTE (USA), UNIVERSITY OF SUSSEX (UK), UNIVERSITY OF HELSINKI (FINLAND), UNIVERSITY COLLEGE LONDON (UK), UNIVERSITY OF QUÉBEC AT MONTRÉAL (CANADA), UNIVERSITY OF BRITISH COLUMBIA

**HIGHLIGHT:** O. GASCUEL, AWARDED OF THE 2017 INRIA - FRENCH ACADÉMIE DES SCIENCES GRAND PRIZE; THE WORK OF M. HEBRARD (POSTDOC NUMEV) RECENTLY LED TO A PUBLICATION ON THE OPEN-ACCESS VERSION OF SCIENCE.

### Objectives

The Integrated Project (IP) “DNA and Genome” concerns the design of novel mathematical and biophysical models to understand genetic and epigenetic processes, and algorithms for the treatment and analysis of the huge amounts of DNA data available today in all life-science domains, as well as in a number of socio-economic activities (health, environment, agriculture, food etc.). By combining biology, mathematics, bioinformatics and physics, this IP is strongly involved in understanding the fundamental principles guiding genome organization, regulation, function and evolution. The IP “DNA and Genomes” was created in 2014, in accordance with the NUMEV Scientific Committee. The idea was to bridge the gap between physical and compu-

tational approaches, which are traditionally developed separately. The areas where the “DNA and Genomes” community are particularly active are: (1) comparative and functional genomics, (2) evolutionary biology, including population genetics, (3) theoretical biological physics of DNA and chromosomes, and (4) the analysis of Next Generation Sequencing data (NGS).

### Contributions

There are a number of close collaborations with biology laboratories in the Montpellier area (notably within the contours of the LabEx EpiGenMed, Cemeb and Agro). These collaborations provided innovative applications and insights into HIV-AIDS and other pathogens, bacterial replication and the biology of micro-organisms

with a role in the carbon cycle, potentially impacting environmental sciences, the management of invasive and harmful populations, health, agriculture, and food science. In the following, we illustrate the major results and actions funded.

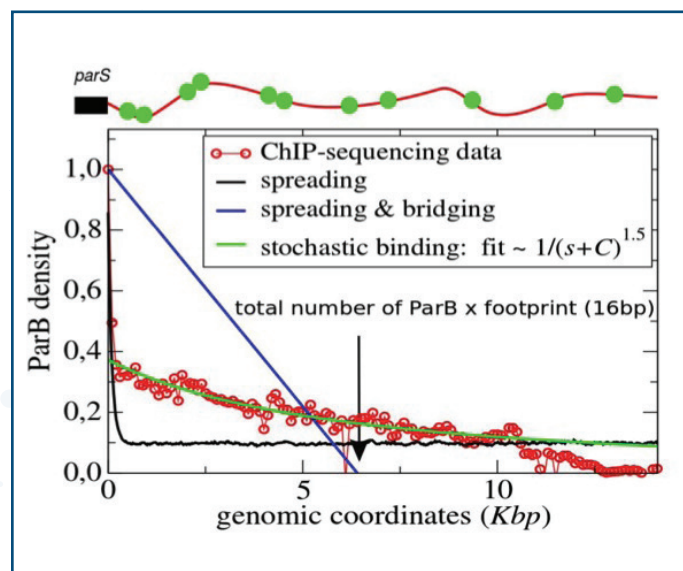
A post-doc was funded to work on the bioinformatics of HIV. A collaboration with the University College London was established to study the transmission of resistance mutations in the UK population, a pressing topic because the spread of these mutations could render inefficient the current drug treatments against HIV. Using an original methodology that extends a first research work<sup>2</sup>, it was shown that a surprisingly large fraction of resistance transmissions take place among drug naïve patients (Aids publication<sup>1</sup>).

In a collaboration with the Banyuls oceanological observatory and the Joint Genome Institute in the USA, the work of Maxime Hebrard (Post-doc) on the genome of the smallest known alga, *O. tauri*<sup>3</sup> recently led to important discoveries on the biology of these pico-algae, showing their sexual reproduction and an intriguing link between genome structure and viral immunity<sup>4</sup>. This work is an interesting example of how bioinformatic analyses can help shed light on the biology of an organism, especially when direct observation is problematic.

Several projects have been carried out in collaboration with the "Algorithms & Computations", notably the PhD Theses of Coralie Merle in population genetics<sup>5</sup> and of Manuel Binet on phylogenomics<sup>6</sup>. Another research work who showed that 3D proximity between genomic regions may play an important role in the structural evolution of genomes<sup>7</sup>. This finding may help the reconstruction of the evolutionary rearrangement scenarios linking two genomes, and opens the way to a number of biologically-relevant algorithmic problems.

With the "Modeling", a leveraging action has been funded to initiate activities in Physical Genomics, with the active participation of the Montpellier biophysics community: two post-docs were funded to support the collaboration of the L2C with the CBS (Montpellier, M. Nollmann) and the LMGM (Toulouse, J.-Y. Bouet) on the physics of active DNA segregation and positioning in bacteria. Physical modeling allows to describe, quantitatively and within a unique framework, the experimental results for the partition complex coming from both high throughout sequencing data (ChIP-sequen-

cing) on protein-DNA interactions and quantitative super-resolution microscopy experiments<sup>8</sup>, see Figure 1. Very recent modeling results also reveal that a non-linear physics approach leads to a coherent description of the dynamics of DNA (partition complex) segregation and stable positioning before cell division in bacteria, a mechanism that has until now resisted full understanding<sup>9</sup>.



**Fig 1: The predictions of three different biophysical models are confronted with the ChIP-sequencing data giving the density of the protein ParB (making up the DNA-protein partition complex) along the DNA chain, starting from the specific *parS* binding site. The Stochastic Binding model (green curve) developed in (Cell Systems 2015 [RI46]) agrees best with the data.**

Together with INSERM, CNRS, LabEx EpiGenMed, Pole BioMED Rabelais, the IP funded a Joint International Meeting "Biophysics across Scales 2015" of the "International Physics of Living Systems" network (GDRI IPoLS, created by CNRS-NSF with major worldwide universities) and the local GRISBi network with a focus on regulatory mechanisms of the genome.

The Mathematical and Computational Evolutionary Biology conferences, supported by NUMEV, continued from 2014 to 2017, with the latest edition entitled «Methods for Integrative Evolutionary Biology: various sources of data, various scales of evolution». This conference, which has become a reference, attracts annually a number of world famous researchers working on these essential questions.

Several scientists with NUMEV funding visited Montpellier for long time periods, including Anne Bergeron,

Professor of bioinformatics in Montreal, working on gene rearrangements (notably in viruses), for 3 months in 2015, a collaboration that led to a publication<sup>10</sup>.

## References

<sup>1</sup> R. Mourad, F. Chevenet, D. T. Dunn, E. Fearnhill, V. Delpéch, D. Asboe, O. Gascuel, S. Hue. (2015) A phylotype-based analysis highlights the role of drug-naïve HIV-positive individuals in the transmission of antiretroviral resistance in the UK. *AIDS*, vol. 29(15), pp.1917-25, doi: 10.1097/QAD.0000000000000768.

<sup>2</sup> F. Chevenet, M. Jung, M. Peeters, T. De Oliveira, O. Gascuel (2013) Searching for Virus Phylotypes. *Bioinformatics*, 29 (5), pp. 561-570

<sup>3</sup> R. Blanc-Mathieu, B. Verhelst, E. Derelle, S. Rombauts, F.-Y. Bouget, I. Carré, A. Château, A. Eyre-Walker, N. Grimsley, H. Moreau, B. Piégu, E. Rivals, W. Schackwitz, Y. Van de Peer, G. Piganeau (2014) An improved genome of the model marine alga *Ostreococcus tauri* unfolds by assessing Illumina de novo assemblies. *BMC Genomics*. Vol. 15 (1103), doi: 10.1186/1471-2164-15-1103 2014

<sup>4</sup> R. Blanc-Mathieu, M. Krasovec, M. Hebrard, S. Yau, E. Desgranges, J. Martin, W. Schackwitz, A. Kuo, G. Salin, C. Donnadieu, Y. Desdevises (2017). Population genomics of picophytoplankton unveils novel chromosome hypervariability. *Science Advances* 1;3(7):e1700239.

<sup>5</sup> C. Merle, R. Leblois, F. Rousset, P. Pudlo (2016), Resampling: an improvement of Important Sampling in varying population size models. *Theoretical Population Biology*, vol. 114, pp. 70-87

<sup>6</sup> M. Binet, O. Gascuel, C. Scornavacca, E. Douzery, F. Pardi (2016) Fast and accurate branch lengths estimation for phylogenomic trees, *BMC Bioinformatics*, *BioMed Central*, 17 (23), doi: <10.1186/s12859-015-0821-8>

<sup>7</sup> S. Pulicani, P. Simonaitis, E. Rivals, K. M. Swenson (2017). Rearrangement scenarios guided by chromatin structure. In *RECOMB International Workshop on Comparative Genomics* (pp. 141-155). Springer

<sup>8</sup> A. Sanchez, D. I. Cattoni, J.-C. Walter, J. Rech, A. Parmeggiani, M. Nollmann, J.-Y. Bouet. (2015) Stochastic Self-Assembly of ParB Proteins Builds the Bacterial DNA Segregation Apparatus. *Cell Systems* Vol. 1 (2), p. 163-173, doi: 10.1016/j.cels.2015.07.013.

<sup>9</sup> J.C. Walter, J. Dornignac, V. Lorman, J. Rech, J.-Y. Bouet, M. Nollmann, J. Palmeri, A. Parmeggiani, F. Geniet (2017). Surfing on protein waves: proteophoresis as a mechanism for bacterial genome partitioning. *Physical Review Letters*, 119(2), 028101

<sup>10</sup> S. Bérard, A. Château, N. Pompidor, P. Guertin, A. Bergeron, K. M. Swenson (2016) Aligning the unalignable: bacteriophage whole genome alignments. *BMC Bioinformatics*, *BioMed Central*, 17 (1), pp. 1-13, doi<10.1186/s12859-015-0869-5>