Organizing committee
Frédéric Brunet (IGFL, Lyon)
Anna-Sophie Fiston-Lavier (ISEM, Montpellier) - François Sabot (IRD, Montpellier)
Emmanuel Talla (LCB, Marseille) - Sébastien Tempel (LCB, Marseille)

Program

Wednesday, July 10, 2019

14H30: Welcome remarks

Session 1 : Algorithms, Tools and DataBases

Chair: Anna-Sophie Fiston Lavier

14H50 – 15H15 - **Emmanuelle Lerat** (LBBE, Lyon) - *Evaluation of tools dedicated to transposable element detection.*

15H15 – 15H40 - **François Sabot** (IRD, Montpellier) - *Methodology for the creation of the panreference and identification of PAV.*

15H40 – 16H05 - **Sebastien Tempel** (LCB, Marseille) - *Dynamics and evolution of transposable elements in procaryotes.*

16H05 – **Coffee break** (25 min)

Session 2 : Large Scale Data Analyses
Chair: François Sabot

16H30 – 16H55 - **Anna-Sophie Fiston-Lavier** (ISEM, Montpellier). *Dynamics of TE insertions in Drosophila re-sequencing data after piwiKd.*

16H55 – 17H20 - **Marie-Christine Carpentier** (LGDP, Perpignan) - *The 3000 rice genomes unravel TE-driven genome dynamics in the field.*

17H20 – 17H45 - **Hadi Quesneville** (URGI, INRA Versailles) - *Traces of past transposable element presence in Brassicaceae genome dark matter.*

19H30 – **Diner**

Thursday, July 11, 2019

Discussion: *Good practices in analyzing Mobile Genetic Elements (MGE)*

9H00-10H30
Session 1. Detection and Annotation of MGE
Moderators: Hadi Quesneville, Jean-Nicolas Volff, Emmanuelle Lerat, and Aurelie Hua Van

10H30- Coffee Break (20 min)

10H50-12H20
Session 2. Large scale analysis of MGE
Moderators: Frédéric Choulet, Marie Fablet, Clémentine Vitte, and Marie-Christine Carpentier

12H20 - Concluding remarks – End

Evaluation of tools dedicated to transposable element detection

Emmanuelle Lerat
Laboratoire de Biométrie et Biologie Evolutive (LBBE), Lyon, France

Since several decades, numerous bioinformatics tools have been developed to identify repeats in genome sequences, especially for the detection of polymorphic insertions. New tools continue to arise regularly to follow the progress in sequencing technology in particular, but also to answer to specific biological questions. Due to their numbers, it becomes difficult to determine which tool is the best to use according to the data and the question. It is thus important to understand and evaluate these tools in particular by testing them on reliable data to benchmark them.

Methodology for the creation of the panreference and identification of PAV

François Sabot

Institut de Recherche pour le Développement (IRD), Montpellier, France

Nowadays, NGS from second and third generation allow any researcher to sequence her/his organism of interest, and even dozens or hundreds of samples in a very short period of times. This provides a raw material of huge interest for TE studies, but also raise the problem of the analyses of such data. Using a non-TE based approach, I will show how we were able to identify Presence/Absence Variation in a reference based approach, as well as in non reference sequences, and what we learned in the evolution of genome structure during domestication of two rice species.

Dynamics and evolution of transposable elements in procaryotes

Sebastien Tempel and Emmanuel Talla

Laboratoire de Chimie Bacterienne (LCB), Marseille, France

Insertion sequences (IS) are small and simple transposable elements that are widely disseminated in bacterial and archaeal genomes. IS are mainly responsible for the mutations and recombinations in prokaryotic genomes and also participate to the gene regulatory network as promoter or transcription factor binding site. From known IS (classified in 29 superfamilies) located in ISFinder webserver and genomic data from NCBI database, we identified of all IS occurrences in prokaryotic genomes. Therefore, IS occurrences were analysed in combination with the genetic context associated to each IS occurrences, as well as regulatory sequences from various databases (such as CollectTF, Prodoric2 or Genome2D).

Altogether, we built a website that presents the impact and putative functional roles of IS in prokaryotic organisms. This website allows the user (1) to dynamically explore the IS occurrences with its genetic environment, (2) to perform novel similarity searches from IS sequences, (3) to display all synthenic IS-Gene couples and the network of horizontal transfer involving the IS, and (4) to predict putative IS role associated to neighboring genes.

Dynamics of TE insertions in Drosophila re-sequencing data after piwiKd

Anna-Sophie Fiston-Lavier

Institut des Sciences de l'Evolution Montpellier (ISEM), Montpellier, France

It has been suggested that, when organisms are facing large environmental or chemical stresses, living beings undergo large metabolic, cellular and genomic modifications that could be due, in part, to genomic elements called transposable elements (TEs). While their implications are of capital relevance, to date, any clues of the repercussions of TEs in evolution are restricted to correlative studies. We lack the possibility to do directed experiments because of two major caveats. First, stress induced experiment may fail. For example, the constitutive loss of TE repression (*i.e.,* piwi-knockdown - piwiKd) leads to sterility. Second, to date the second generation of sequencing technology is not adapted for detecting accurately novel insertions with 40% of false-positive. Here, I will present an approach to date TE insertions over generations after piwiKd in *Drosophila melanogaster*.

Unraveling the Retrotranspositional Landscape of Rice at Species Level using 3000 Genomes

Marie-Christine Carpentier[1], Ernandes Manfroi[2], Fu-Jin Wei[3], Hshin-Ping Wu[3,4], Eric Lasserre[1], Christel Llauro[1], Emilie Debladis[1], Roland Akakpo[1], Yue-Ie Hsing[3] & Olivier Panaud[1,5]

1- Laboratoire Génome et Développement des Plantes, UMR CNRS/UPVD 5096. Université de Perpignan Via Domitia. 52 avenue Paul Alduy. 66860 Perpignan cedex. France. 2- Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul. Brazil. 3- Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan 11529. 4- Department of Forest Molecular Genetics and Biotechnology, Forestry and Forest Products Research Institute, Tsukuba, Japan. 5- Institut Universitaire de France. 1, rue Descartes. 75231 Paris cedex 05.

Rice, Oryza sativa, is the staple food for half the world population. It is the first crop the genome of which was sequenced ten years ago (IRGSP 2005). The availability of large genomic datasets made possible by the cost-effectiveness of next generation sequencing (NGS) provides the opportunity to study genome evolution at the species level at an unprecedented scale. In rice, the release of such data for 3,000 varieties is a breakthrough in crop genomics, giving access to the genetic diversity of the whole cultivated gene pool in a single dataset. We took advantage of this resource to unravel the retrotranspositional landscape of rice. We developed a new software, TRACKPOSON, specifically designed for the detection of transposable elements insertion polymorphisms (TIPs) from large datasets and applied it to 31 families of retrotransposons. Our results show that most polymorphisms are found at very low frequencies in rice germplasm, which suggests that they occurred recently, in agro. Using a genome-wide association studies, we also show that the activation of transposition in rice is probably triggered by an external stimulus, rather than by the impediment of a silencing pathway. We also used the TIPs dataset to trace the origin of rice domestication and confirm that rice originated from two distinct domestication events.

Traces of past transposable element presence in *Brassicaceae* genome dark matter

Agnès Baud[1], Mariène Wan[1], Danielle Nouaud[2], Dominique Anxolabéhère[2], Hadi Quesneville[1]

[1]*URGI, INRA, Université Paris-Saclay, 78026, Versailles, France*

[2]*IJM, Institut Jacques Monod, CNRS UMR 7592, Université Paris-Diderot, Paris, France*

Transposable elements (TEs) are mobile, repetitive DNA sequences that have been found in every branch of life. In many organisms TEs are the primary contributors to the genome bulk. They invade genomes recurrently by wave of transposition bursts that ceased rapidly as repressed by host defense mechanisms subsequently triggered. The sequences become immobile and start to degrade, fading away in the genome sequence so that it cannot be recognized as such. It contributes then to the so-called "dark matter of the genome", this part of the genome where nothing can be recognized as biologically functional in first instance.

We developed a new method able to find these old and degenerated TE sequences. With the new algorithm we implemented, we detect up to 10% of the *A. thaliana* genome deriving from TEs not yet identified. Altogether we bring to 50% the part of the genome deriving from TE in this species. Interestingly these sequences are generally very short, about 500bp, and found in the upstream 500pb of genes. Their epigenetic status and their nucleotide composition suggest an old TE origin. A significant fraction of them overlap conserved non-coding sequences identified in crucifers and rosids, and transcription factor binding sites. They are over-represented in the flowering gene regulation network. Altogether, this suggests a functional role of these sequences being conserved over 100 million years, since the Cretaceous when flowering plants appears.