

# Data Structures in Bioinformatics

## DSB 2024 - detailed program

DSB 2024 team

March 8, 2024

### Location and dates

DSB 2024 takes place in Montpellier, South of France  
on the Campus St Priest, from Montpellier University  
Amphitheater "Jean-Jacques Moreau", Building #2  
on March 14-15 in 2024.  
Lunches and social events are located in Building #5.

### Organizers

From team "Methods & Algorithms for Bioinformatics":

- Jordan Moutet (LIRMM - CNRS et Univ. Montpellier)
- Eric Rivals (LIRMM - CNRS et Univ. Montpellier)
- Nikolai Romashchenko (LIRMM - CNRS et Univ. Montpellier)
- Pengfei Wang (LIRMM - CNRS et Univ. Montpellier)

and the LIRMM Communication unit

- Virginie Fêche
- Elena Demchenko

### Sponsors

We are grateful to the sponsors of the 2024 edition:

- The Int'l Training Network (ITN) **ALPACA**, European Union's Horizon 2020 research and innovation

programme under the Marie Skłodowska-Curie grant agreement No 956229.



- The **GDR Bioinformatique Moléculaire** from CNRS



## Day 1 Thursday 14th March

- 8h45 Welcome address
- Session 1 - 9:00-10:30
  - 9:00-09:30 Chikhi Rayan - *Work in progress for peta-scale sequence exploration*
  - 9:30-10:00 Depuydt Lore - *r-indexing without backward searching*
  - 10:00-10:30 Caleb Smith - *Brisk: Exact resource-efficient dictionary for k-mers*
- break
- Session 2 - 11:15-12:30
  - 11:15-11:45 Draesslerová Dominika - *Metagenomic classification with maximal exact matches in KATKA kernels and minimizer digests*
  - 11:45-12:15 Romashchenko Nikolai - *Memory-frugal disk-based (phylo-)k-mer filtering for alignment-free phylogenetic placement*
  - 12:15-12:30 Rouzé Timothé - *Pangenomic k-mer distribution with low memory cost*
- 12:30-14:00 lunch
- Session 3 - 14:00-15:30
  - 14:00-14:30 Avila Jorge - *A compact embedding-based indexing for accurate and rapid classification in bacterial pangenomics*
  - 14:00-14:30 Rizzo Nicola - *Exploiting uniqueness: seed-chain-extend alignment with Elastic Founder Graphs*
  - 15:00-15:30 Levallois Victor - *The Backpack Quotient Filter: a dynamic and space-efficient data structure for querying k-mers with abundance*
- break
- Session 4 - 16:15-18:00
  - 16:15-16:45 Luca Parmigiani - *Interpolating and Extrapolating Node Counts in Colored Compacted de Bruijn Graphs for Pangenome Growth Comparison*
  - 16:45-17:15 Bonnet Konstantinn - *Strangeppg: Toward Pangenome Scale Graph Visualization*
  - 17:15-17:45 Dubois Siegfried - *Towards an edit distance between pangenome graphs*
- 18:15-21:00 Social event "Food and Wine"

## Day 2 Friday 15th March

- Session 5 - 9:00-10:00
  - 09:00-09:30 Rivals Eric - *An incremental algorithm for computing the set of all period sets*
  - 9:30-10:00 Renders Luca - *Automated design of efficient search schemes for lossless approximate pattern matching*
- break
- Session 6 - 10:45-12:15
  - 10:45-11:15 Hendrychová Veronika - *Mathematical model of phylogenetic compression*
  - 11:15-11:45 Wirtz Johannes - *Counting multiple-merger tree structures emerging in Population Genetics*
  - 11:45-12:15 Sgro Mattia - *PlasBin-flow on Pangenome graphs: improving bacterial plasmid binning in short-read assemblies*
- 12:15-14:00 lunch
- Session 7 14:00-15:30
  - 14:00-14:30 Vandamme Lea - *Tinted de Bruijn Graphs for efficient read extraction from sequencing datasets*
  - 14:30-14:45 Martayan Igor - *Conway-Bromage-Lyndon (CBL): an exact, dynamic representation of  $k$ -mer sets*
  - 14:45-15:30 Biagi Elena - *Finimizers: Variable-length bounded-frequency minimizers for  $k$ -mer sets*
- break
- Session 8 - 16:15-17:15
  - 16:15-16:45 Schmitz Johanna - *EpiSegMix: Discovering chromatin states using a flexible distribution hidden Markov model with duration modeling*
  - 16:45-17:15 Uzuner Hamdiye - *Generalized uncertainty-aware haplotype quantification with application in HLA typing and virus analysis*
- 17:15 Closing