



Session de formation 2023



bioinformatics platform dedicated to the genetics and genomics of tropical and Mediterranean plants and their pathogens

génomique formations ressources Infrastructure montpelliérain
plantes internationale orienté développement
sud service calcul développement
Reseau plateforme d'analyses
compétences végétale multi-instituts
communauté outils mutualisation partage
s'appuie cassava mutualisation partage



SNP detection genome assembly
phylogeny transcriptome assembly differential expression
comparative genomics structural variation
GWAS pangenomics
population genetics polypliody metapopulation

Mutualisation



Cacao

Banana

Coffee

Rice

Palm

Cassava

Pseudocercospora

Magnaporthe

South Green

bioinformatics platform



4 institutes



25+



3 research units



Tools

Storage and computing
resources



400+

Trainings



Meso@LR au CINES

1090 threads :

35 standard nodes

2 bigmem nodes

1 GPU node

500 To of replicated storage



CINES

1130 threads:

30 standard node

1 supermem node

1 GPU node

150 To on 3 NAS + 210 To scratch



400+



600+ tools



Resources mutualised at Meso@LR through the
Mudis4Ls project (purchase/storage/data)

Collaborative development of tools

Genomics

Pangenomic

Gene families

Comparative

Phylogeny

Assemblies

Annotation

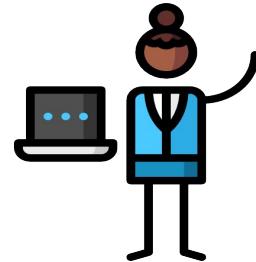
Data mining

Diversity exploration

genotype manipulation

mosaic manipulation

Metagenomic



+20
tools

web applications (16)

visualisation (8)

workflows(5)

packages (4)



<https://github.com/SouthGreenPlatform/>



Plant & Health Bioinformatics Platform



<https://bioinfo.ird.fr/>



AURORE
COMTE



JACQUES
DAINAT



ALEXIS
DEREOPER



BRUNO
GRANOUILLAC



JULIE
ORJUELA-



NDOMASSI
TANDO



CHRISTINE
TRANCHANT



bioinfo@ird.fr



@ItropBioinfo



Florian Charriat
Antoni Exbrayat



Guilhem Sempere



Bruno Granouillac
Jacques Dainat



Nicolas Fernandez

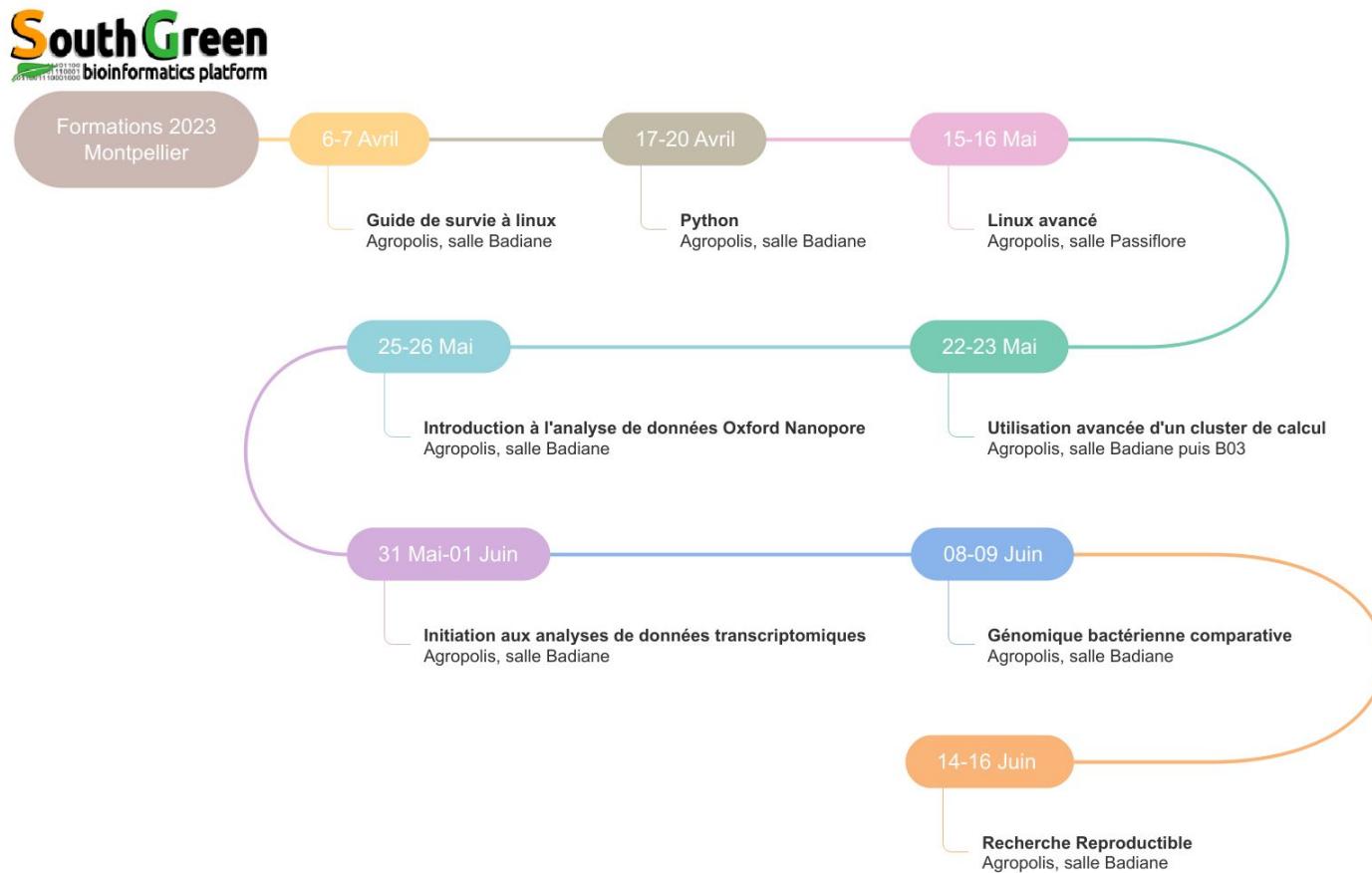


Thomas Denecker

And more collaborators !

South Green

bioinformatics platform



Modules de formation 2023

- Toutes nos formations :
<https://southgreenplatform.github.io/trainings/>
- Topo & TP :
https://github.com/SouthGreenPlatform/training_ONT_teaching/tree/2023_MTP
- Environnement de travail : [Logiciels à installer](#)



Génomique Comparative Bactérienne



Two Approaches to Microbial Genomics

Starting with sets of reads representing your study isolates...



Assembly-based

1. Assemble each set of reads into a genome sequence
2. Annotate each genome
3. Cluster genes and compare between each genome

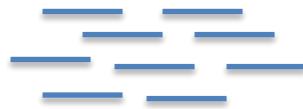
Variant-based

1. Compare each read set to a reference genome assembly
2. Directly compare variants between each genome

Two Approaches to Microbial Genomics

Starting with sets of reads representing your study isolates...

A



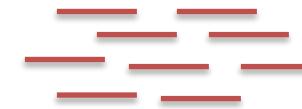
B



C



D



Assembly-based

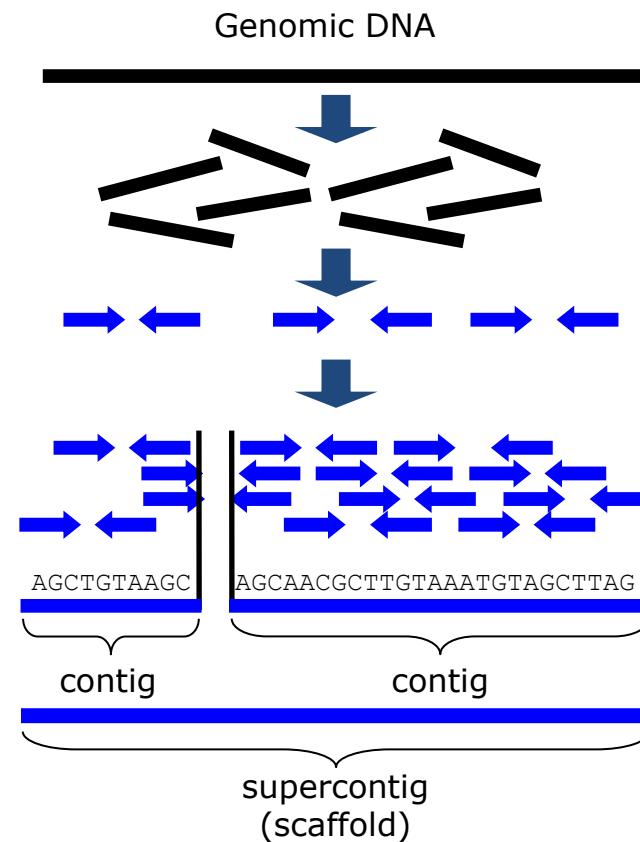
1. Assemble each set of reads into a genome sequence
2. Annotate each genome
3. Cluster genes and compare between each genome

Variant-based

1. Compare each read set to a reference genome assembly
2. Directly compare variants between each genome

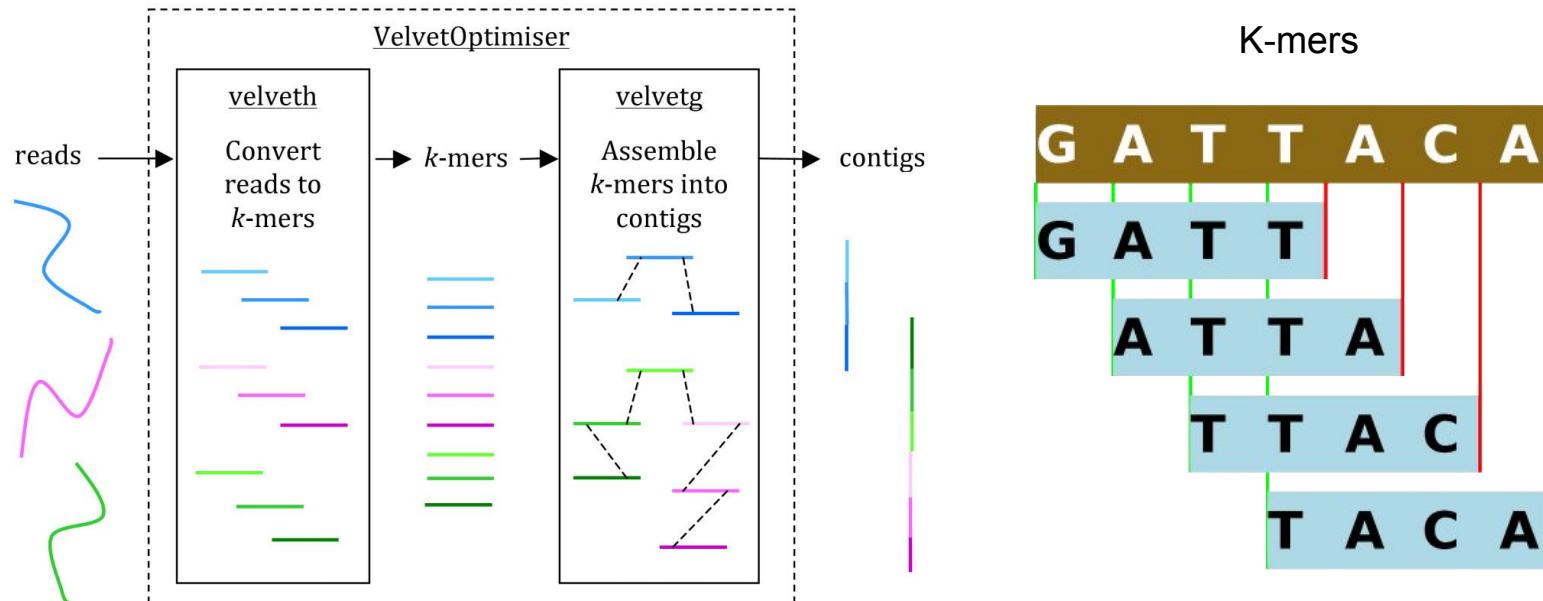
1) Assembly

Assembly Basics (de-novo assembly)



Assembly Methods

- SPAdes (<http://cab.spbu.ru/software/spades/>)
- Velvet (<https://www.ebi.ac.uk/~zerbino/velvet/>)
- Both are De Bruijn graph assemblers





Brief Report

Comparison of De Novo Assembly Strategies for Bacterial Genomes

Pengfei Zhang^{1,2,†}, Dike Jiang^{1,2,†}, Yin Wang^{1,2,*}, Xueping Yao^{1,2}, Yan Luo^{1,2} and Zexiao Yang^{1,2}

Table 1

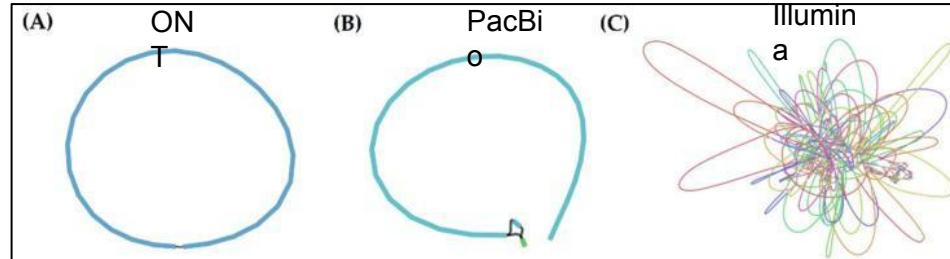
Statistics of genome-assembly results of independent assembly strategies.

Platforms	Assembler	Contigs	Largest Contig (bp)	N50	GC%
Illumina	SPAdes	527	157,573	40,498	39.87
PacBio	Canu	25	2,351,556	2,351,556	40.01
ONT	Canu	1	2,360,091	2,360,091	40.02

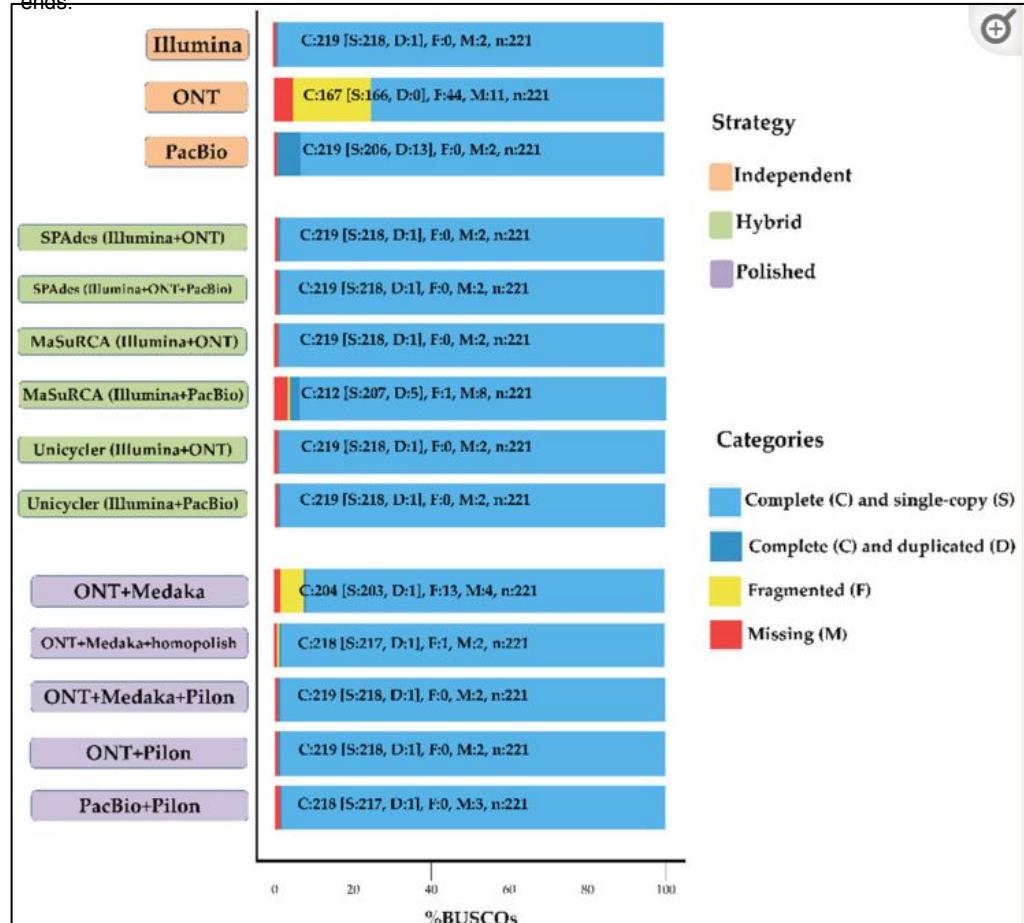
Table 2

Statistics of genome-assembly results of hybrid assembly strategies.

Platforms	Assembler	Contigs	Total Length (bp)	N50	GC%
Illumina + ONT	SPAdes	266	2,402,219	1,953,224	39.97
Illumina + PacBio + ONT	SPAdes	236	2,410,042	2,351,543	40.02
Illumina + ONT	Unicycler	1	2,349,186	2,349,186	40.03
Illumina + PacBio	Unicycler	1	2,349,340	2,349,340	40.03
Illumina + ONT	MaSuRCA	1	2,365,339	2,365,339	40.02
Illumina + PacBio	MaSuRCA	4	2,395,409	1,345,876	40.04

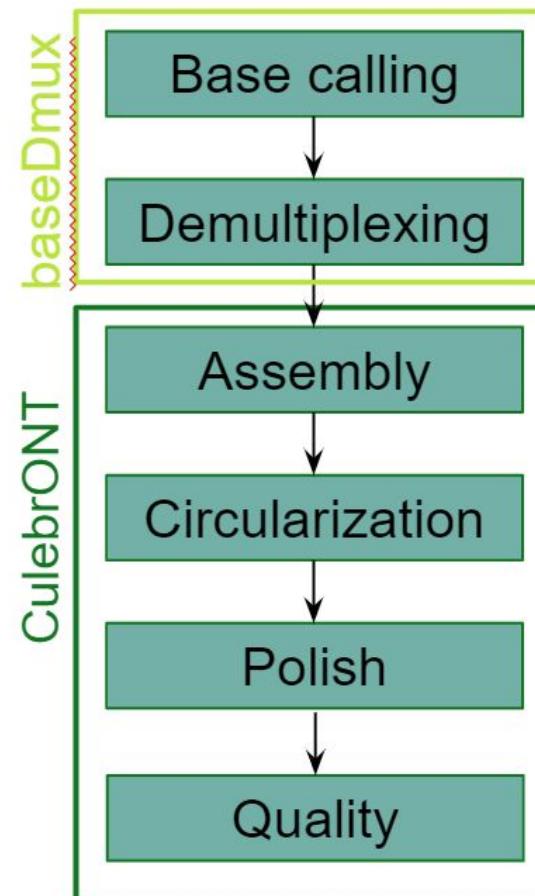
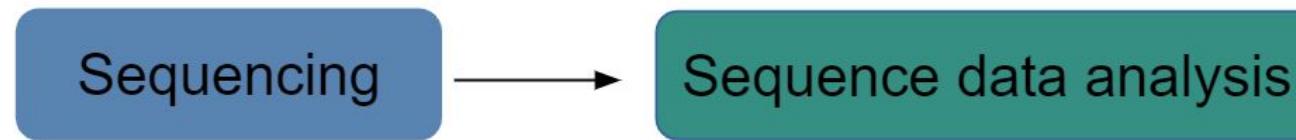


Comparison of results of independent assembly strategies. (A) Genome assembled with nanopore reads; (B) longest contig assembled with PacBio reads; (C) genome assembled with Illumina reads. Plots were obtained by using Bandage on the “assembly_graph.gfa” output file from SPAdes or the “contig.gfa” output file from Canu. Connections between contigs represent overlaps between contig ends.



Evaluation of completeness of assembly results of different strategies. Assessments of the completeness of the assembly genomes with the datasets of proteobacteria_odb9 lineage. Bar charts produced with BUSCO plotting tool to show proportions that were classified as complete (C, blue), complete single copy (S, light blue), complete duplicated (D, dark blue), fragmented (F, yellow), and missing (M, red).

Bioinformatic Workflows: assembly



Snakemake



<https://github.com/vibaotram/baseDmux>



<https://culebront-pipeline.readthedocs.io/en/latest/>



2) Separate chromosomal and plasmid
scaffolds/contigs

MOB-suite: Software tools for clustering, reconstruction and typing of plasmids from draft assemblies

Introduction

Plasmids are mobile genetic elements (MGEs), which allow for rapid evolution and adaption of bacteria to new niches through horizontal transmission of novel traits to different genetic backgrounds. The MOB-suite is designed to be a modular set of tools for the typing and reconstruction of plasmid sequences from WGS assemblies.

The MOB-suite depends on a series of databases which are too large to be hosted in git-hub. They can be downloaded or updated by running `mob_init` or if running any of the tools for the first time, the databases will download and initialize automatically if you do not specify an alternate database location. However, they are quite large so the first run will take a long time depending on your connection and speed of your computer. Databases can be manually downloaded from [here](#).

Our new automatic chromosome depletion feature in MOB-recon can be based on any collection of closed chromosome sequences.

Citations

Below are the manuscripts describing the algorithmic approaches used in the MOB-suite.

1. Robertson, James, and John H E Nash. "MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies." *Microbial genomics* vol. 4,8 (2018): e000206. doi:10.1099/mgen.0.000206
2. Robertson, James et al. "Universal whole-sequence-based plasmid typing and its utility to prediction of host range and epidemiological surveillance." *Microbial genomics* vol. 6,10 (2020): mgen000435. doi:10.1099/mgen.0.000435

MOB-init

On first run of MOB-typer or MOB-recon, MOB-init (invoked by `mob_init` command) should run to download the databases from figshare, sketch the databases and setup the blast databases. However, it can be run manually if the databases need to be re-initialized OR if you want to initialize the databases in an alternative directory.

MOB-cluster

This tool creates plasmid similarity groups using fast genomic distance estimation using Mash. Plasmids are grouped into clusters using complete-linkage clustering and the cluster code accessions provided by the tool provide an approximation of operational taxonomic units OTU's. The plasmid nomenclature is designed to group highly similar plasmids together which are unlikely to have multiple representatives within a single cell and have a strong concordance with replicon and relaxase typing but is universally applicable since it uses the complete sequence of the plasmid itself rather than specific biomarkers.

MOB-recon

This tool reconstructs individual plasmid sequences from draft genome assemblies using the clustered plasmid reference databases provided by MOB-cluster. It will also automatically provide the full typing information provided by MOB-typer. It optionally can use a chromosome depletion strategy based on closed genomes or user supplied filter of sequences to ignore.

MOB-typer

Provides *in silico* predictions of the replicon family, relaxase type, mate-pair formation type and predicted transferability of the plasmid. Using a combination of biomarkers and MOB-cluster codes, it will also provide an observed host-range of your plasmid based on its replicon, relaxase and cluster assignment. This is combined with information mined from the literature to provide a prediction of the taxonomic rank at which the plasmid is likely to be stably maintained but it does not provide source attribution predictions.

MICROBIAL GENOMICS

METHODS PAPER

Robertson and Nash, *Microbial Genomics* 2018;4
DOI 10.1099/mgen.0.000206



MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies

James Robertson¹ and John H. E. Nash^{2,*}

3) Genome Annotation

What is annotation ?

Structural annotation:



Find out where the regions of interest (usually genes) are in the sequence data and what they look like.

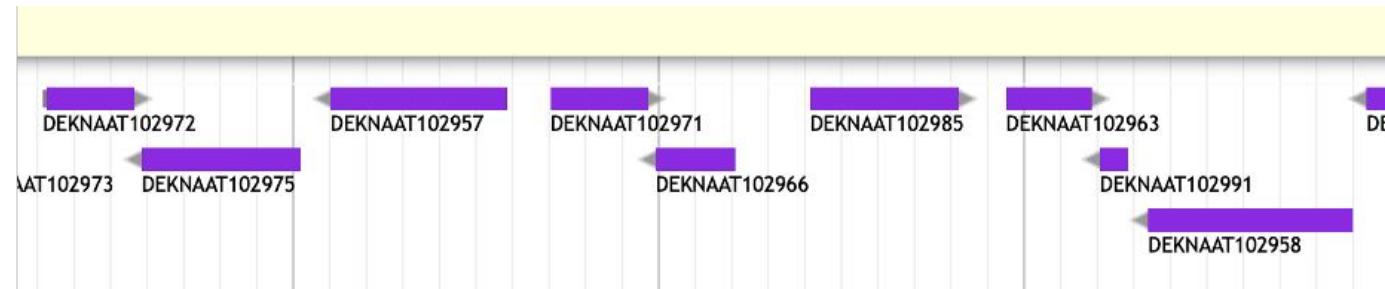
Functional annotation:

Find out what the regions do. What do they code for?

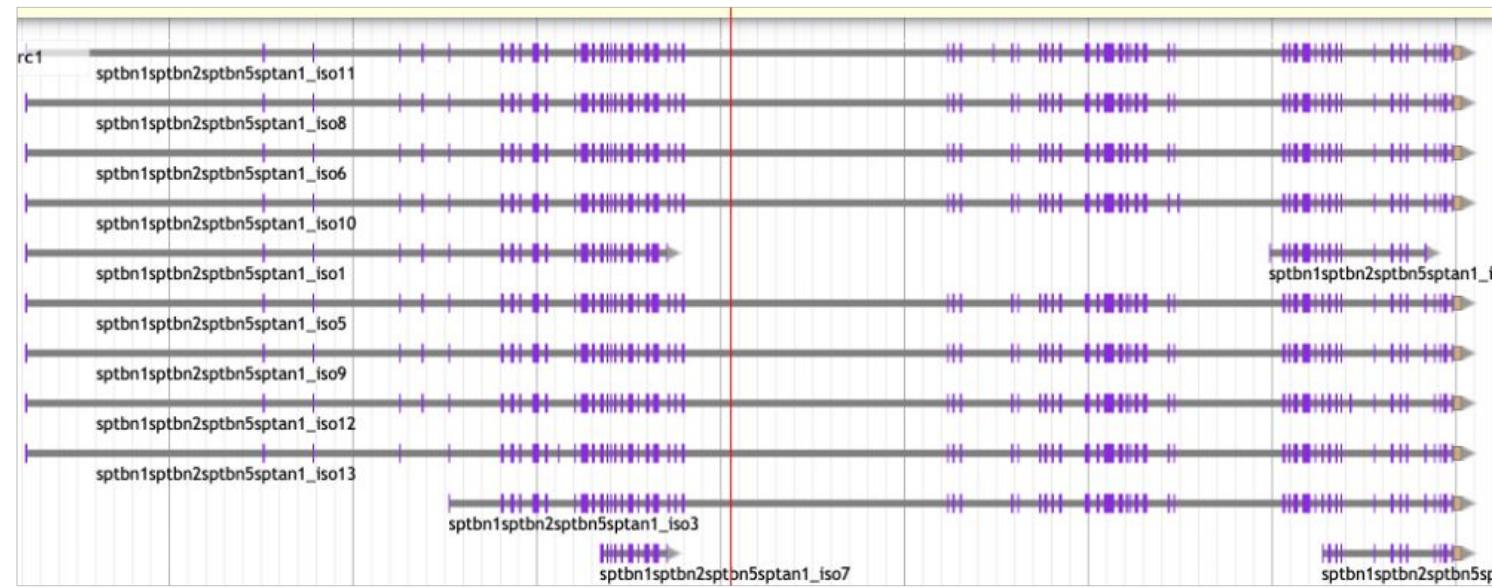
*It is the **annotation** that bridges the gap from the sequence to the biology of the organism*

Organisms differ in genomic complexity

A yeast



A crustacean



##gff-version 3.2.1

##sequence-region ctg123 1 1497228

Header

9 columns

1 feature = 1 line

Ctg123	.	Gene	1000	9000	.	+	.	ID=gene1;Name=EDEN
ctg123	.	mRNA	1050	9000	.	+	.	ID=mRNA1;Parent=gene1;Name=EDEN.1
ctg123	.	mRNA	1050	9000	.	+	.	ID=mRNA2;Parent=gene1;Name=EDEN.2
ctg123	.	exon	1300	1500	.	+	.	ID=exon1;Parent=mRNA3
ctg123	.	exon	1050	1500	.	+	.	ID=exon2;Parent=mRNA1,mRNA2
ctg123	.	exon	3000	3902	.	+	.	ID=exon3;Parent=mRNA1
ctg123	.	exon	5000	5500	.	+	.	ID=exon4;Parent=mRNA1,mRNA2
ctg123	.	exon	7000	9000	.	+	.	ID=exon5;Parent=mRNA1,mRNA2
ctg123	.	CDS	1201	1500	.	+	0	ID=cds1;Parent=mRNA1;Name=eden1
ctg123	.	CDS	3000	3902	.	+	0	ID=cds1;Parent=mRNA1;Name=eden1
ctg123	.	CDS	5000	5500	.	+	0	ID=cds1;Parent=mRNA1;Name=eden1
ctg123	.	CDS	7000	7600	.	+	0	ID=cds1;Parent=mRNA1;Name=eden1
Ctg123	.	CDS	1201	1500	.	+	0	ID=cds2;Parent=mRNA2;Name=eden2
ctg123	.	CDS	5000	5500	.	+	0	ID=cds2;Parent=mRNA2;Name=eden2
Ctg123	.	CDS	7000	7600	.	+	0	ID=cds2;Parent=mRNA2;Name=eden2

- 1) sequence id
- 2) source
- 3) feature type
- 4) start
- 5) end
- 6) score
- 7) strand
- 8) phase

(SO term = 2278 possibilities)

9) attributes
tag=value

! Features are grouped by **parent** relationship

Adding biological info to sequences

ribosome
binding site

delta toxin
PubMed: 15353161

ACCGGCCGAGACA GCGAGCATATGCAGGAAGCGGCAGGAATAAGGA
AAAGCAGCCTCCTGACTTCCCTCGCTTGGTGGTTGAGTGGACCTC
CCAGGCCAGTGCCGGGCCCCCTCATAGGAGAGGAAGCTCGGGAGGTG
GCCAGGGCGCAGGAAGGCGCACCCCCCCCAGCAATCCGCGCGCCGGG
ACAGAATGCCCTGCAGGAATTCTTAGAACAGACCTTCCTCCTG
CAAATAAAACCTCACCCATGAATGCTCACGCAAGTTAATTACAGA
CCTGAAACAAGATGCCATTGTCCCCCGGCCTCCTGCTGCTGCT
CTCCGTCCGTCCGTGGGCCACGGCCACCGCTTTTTTTTGTGCC

transfer RNA
Leu-(UUR)

tandem repeat
CCGT x 3

homopolymer
10 x T

Annotation Methods

- There are different annotation algorithms for protein-coding genes, tRNAs, rRNAs, other non-coding RNAs
- Pipelines exist for performing several in one go

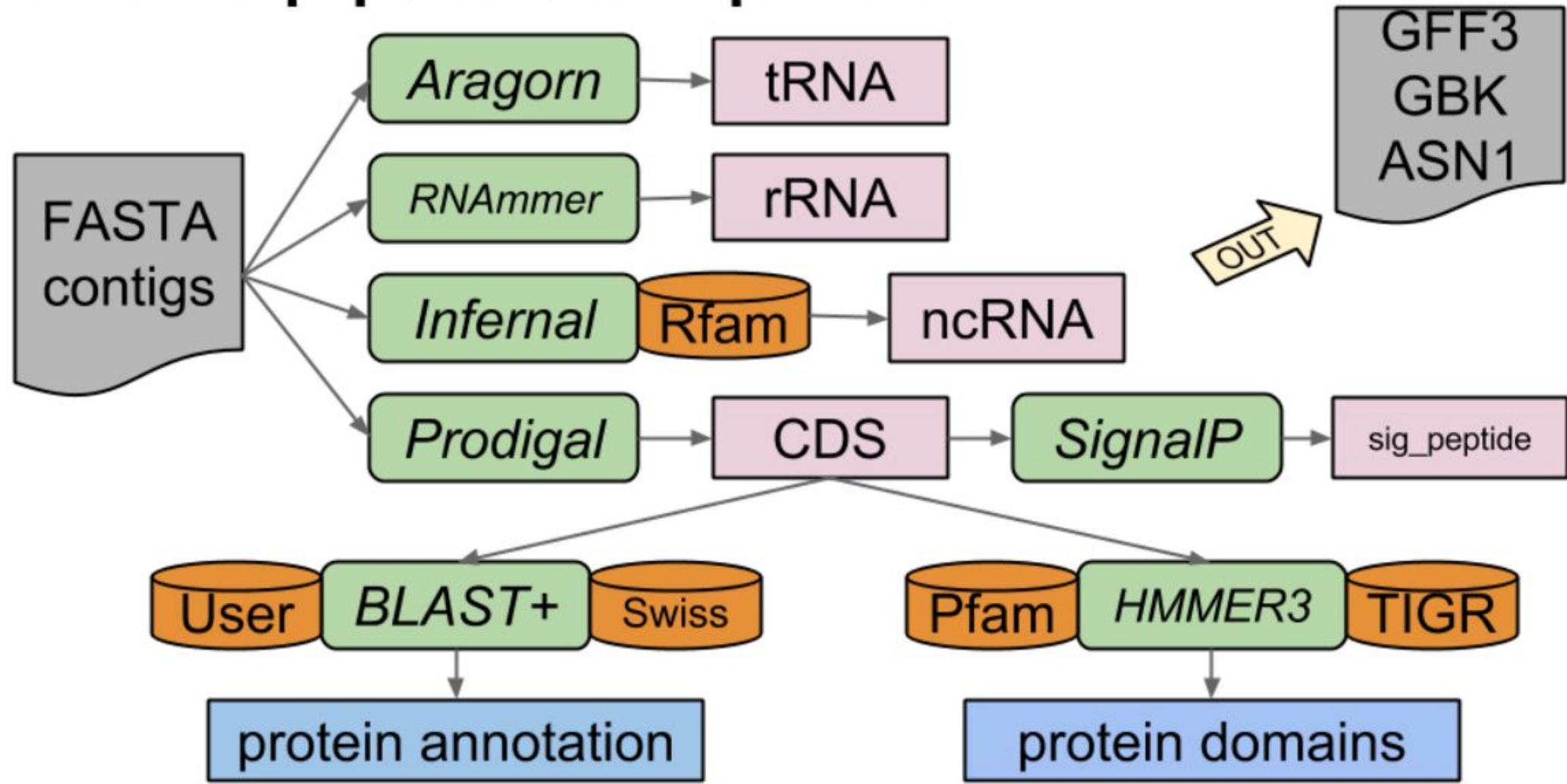
Prokaryote annotation:

- Prokka
(<http://www.vicbioinformatics.com/software.prokka.shtml>) is an all-in-one wrapper for these tools

Table 1. Feature prediction tools used by Prokka

Tool (reference)	Features predicted
Prodigal (Hyatt 2010)	Coding sequence (CDS)
RNAmmer (Lagesen <i>et al.</i> , 2007)	Ribosomal RNA genes (rRNA)
Aragorn (Laslett and Canback, 2004)	Transfer RNA genes
SignalP (Petersen <i>et al.</i> , 2011)	Signal leader peptides
Infernal (Kolbe and Eddy, 2011)	Non-coding RNA

Prokka pipeline (simplified)



Prokaryote annotation:

- Bakta: rapid & standardized annotation of bacterial genomes, MAGs & plasmids
(<https://github.com/oschwengers/bakta>)

Schwengers O., Jelonek L., Dieckmann M. A., Beyvers S., Blom J., Goesmann A. (2021). Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microbial Genomics*, 7(11). <https://doi.org/10.1099/mgen.0.000685>

Tools

- tRNAscan-SE
- Aragorn
- INFERNAL
- PILER-CR
- Prodigal
- Hmmer
- Diamond
- Blast+
- AMRFinderPlus
- DeepSig

Databases

- Rfam
- DoriC: AntiFam
- UniProt
- RefSeq
- COG
- KEGG
- PHROG
- AMRFinder
- ISFinder
- Pfam
- VFDB

4) Public genomes retrieval

National Library of Medicine
National Center for Biotechnology Information

Search NCBI Search

Genomes – NCBI Datasets BETA

Download a genome dataset including genome, transcript and protein sequence, annotation and a data report

TAXONOMIC NAME

Q Anaplasmataceae 1

ASSEMBLY LEVEL

contig scaffold chromosome 2 complete

YEAR RELEASED

1980 3

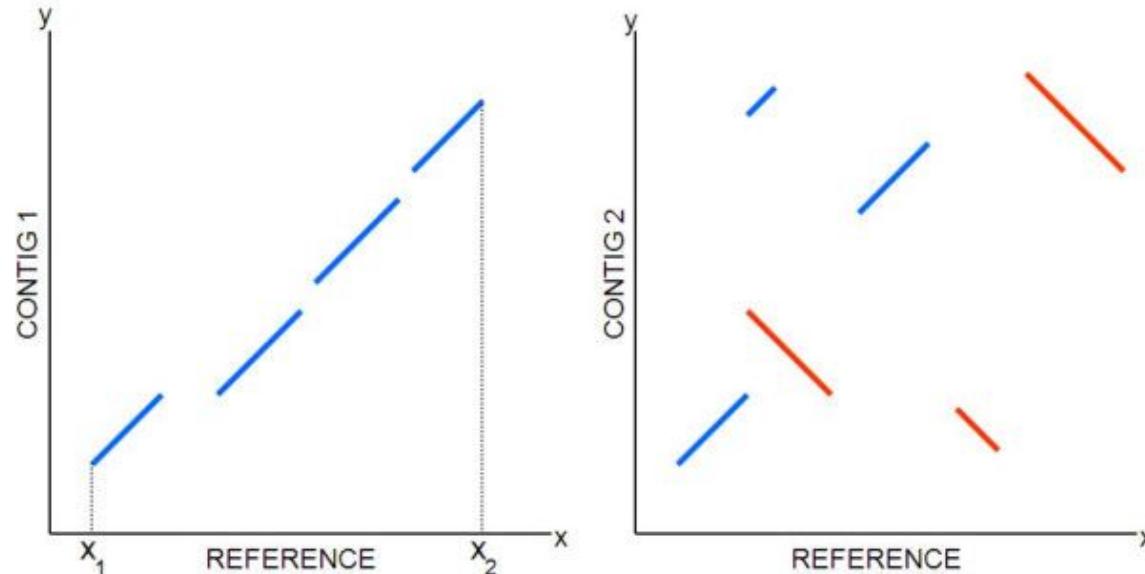
TEXT FILTER

Download 4

Assembly Name	Assembly Accession	Organism	Assembly Level	Submission Date
ASM802v1	GCA_000008025.1	Wolbachia endosymbiont -wMeli	Annotation	12/07/2004
ASM1194v1	GCA_000011945.1	Anaplasma marginale str. Ticks, Mates	Annotation	12/04/2004
ASM1209v1	GCA_000012095.1	Anaplasma marginale str. Ixodes, Mates	Annotation	12/04/2004
ASM1321v3	GCA_000013215.1	Anaplasma phagocytophila, Ix2	Annotation	12/08/2005
ASM1331v2	GCA_000013315.1	Anaplasma phagocytophila, Ix2	Annotation	12/02/2006
ASM1338v1	GCA_000013385.1	Neorickettsia sennetsu str. Miyayama	Annotation	12/02/2006
ASM2030v1	GCA_000020305.1	Anaplasma marginale str. Florida	Annotation	03/12/2009
ASM2228v1	GCA_000022285.1	Wolbachia sp. wili	Annotation	04/01/2009
ASM2235v2	GCA_000022355.1	Neorickettsia risticii str. illinois	Annotation	03/07/2009
ASM2405v1	GCA_000024055.1	Anaplasma centrale str. Isr Israel	Annotation	12/08/2006
ASM2600v1	GCA_000026005.1	Ehrlichia ruminantium str. Weigelevoorde	Annotation	05/12/2005
ASM4954v1	GCA_000049545.1	Ehrlichia ruminantium str. Gardel	Annotation	01/02/2005
ASM4955v1	GCA_000049555.1	Wolbachia ruminantium str. Weigelevoorde	Annotation	01/02/2005
ASM4956v1	GCA_000049565.1	Wolbachia endosymbiont -wRi	Annotation	11/04/2005
ASM4958v1	GCA_000049585.1	Wolbachia endosymbiont -wOr	Annotation	09/17/2012
ASM4959v1	GCA_000049595.1	Wolbachia endosymbiont -wNo	Annotation	12/01/2013
ASM4960v1	GCA_000049605.1	Wolbachia endosymbiont -wHa	Annotation	12/04/2013
ASM4975v1	GCA_000499755.1	Anaplasma phagocytophila, Ix22	NCBI Prokary	04/07/2013
ASM4977v1	GCA_000499775.1	Anaplasma phagocytophila, JM	NCBI Prokary	04/07/2013
ASM4984v1	GCA_000499845.1	Anaplasma marginale str. Gypsy Plains	NCBI Prokary	12/08/2013
ASM4985v1	GCA_000499855.1	Anaplasma marginale str. Dawn	NCBI Prokary	05/12/2013
ASM4986v1	GCA_000499865.1	Ehrlichia ruminantium str. AS45	NCBI Prokary	12/08/2013
ASM4987v1	GCA_000499875.1	Wolbachia endosymbiont -wHarland	Annotation	12/17/2013
ASM4988v1	GCA_000499885.1	Wolbachia sp. sp. wII	Annotation	12/17/2013
ASM4989v1	GCA_000499895.1	Wolbachia endosymbiont str. Ju-Jee	Annotation	12/17/2014
ASM4990v1	GCA_000499905.1	Ehrlichia chaffeensis str. L.Liberty	Annotation	12/17/2014
ASM4991v1	GCA_000499915.1	Wolbachia endosymbiont str. O-Osculata	Annotation	12/17/2014
ASM4992v1	GCA_000499925.1	Ehrlichia chaffeensis str. S.Saint Vincent	Annotation	12/17/2014
ASM4993v1	GCA_000499935.1	Wolbachia endosymbiont str. W.Wakulla	Annotation	12/17/2014
ASM4994v1	GCA_000499945.1	Wolbachia chaffeensis str. W.West Paces	Annotation	12/17/2014
ASM4995v1	GCA_000499955.1	Neorickettsia helminthiae, Oregon	Annotation	12/17/2014
ASM4996v2	GCA_000499965.2	Anaplasma phagocytophila, Norway variant2	NCBI Prokary	03/05/2016
ASM4997v1	GCA_000499975.1	Wolbachia endosymbiont -wCe	Annotation	12/17/2014
Wv-0003	GCA_000499985.1	Anaplasma phagocytophila, Wolbachia simillima wka	Annotation	12/17/2014
WTFRI_1_B	GCA_001499985.1	Wolbachia endosymbiont -wTPE	Annotation	11/09/2016
ASM17269v1	GCA_001726945.1	Wolbachia endosymbiont -wMeli_Cu	NCBI Prokary	12/07/2016
ASM17280v1	GCA_001728045.1	Wolbachia endosymbiont -wMeli_SM	NCBI Prokary	12/07/2016
ASM19317v2	GCA_001931755.2	Wolbachia endosymbiont -Berlin	NCBI Prokary	05/06/2018
ASM22146v2	GCA_002214625.2	Anaplasma ovis str. Halber Heilbr	NCBI Prokary	05/07/2018
ASM22748v2	GCA_002274845.2	Wolbachia pipiens -wAB-BN2016	NCBI Prokary	04/08/2018
ASM22791v2	GCA_002279145.2	Wolbachia pipiens -wAB-FL2005	NCBI Prokary	04/07/2018
ASM26799v1	GCA_002679995.1	Ehrlichia canis, Y2-1	NCBI Prokary	12/02/2018
ASM3151575v1	GCA_003151575.1	Anaplasma marginale, Palmeira	NCBI Prokary	12/09/2018
ASM3151576v1	GCA_003151576.1	Anaplasma marginale, Palmeira	NCBI Prokary	12/09/2018
ASM39995v1	GCA_003999555.1	Neorickettsia endemicum, China	NCBI Prokary	08/09/2018
ASM401712v1	GCA_004017125.1	Wolbachia pipiens wMeli wAb	NCBI Prokary	12/04/2018
ASM49553v1	GCA_004955355.1	Wolbachia endosymbiont of Brugia malayi	NCBI Prokary	16/04/2019
ASM49559v1	GCA_004955955.1	Wolbachia endosymbiont -wMau	NCBI Prokary	12/15/2017
ASM49759v1	GCA_004975955.1	Wolbachia endosymbiont -wMau	NCBI Prokary	12/15/2017
ASM54229v1	GCA_005422955.1	Wolbachia endosymbiont of Carpocoris satasai	wCeuA	04/07/2019
ASM79726v1	GCA_007972685.1	Wolbachia pipiens -wMet_N2S	NCBI Prokary	12/06/2019
ASM79729v1	GCA_007972995.1	Wolbachia pipiens -wMet_O2	NCBI Prokary	12/06/2019
ASM79729v1	GCA_007972995.1	Wolbachia pipiens -wMet_D26	NCBI Prokary	12/06/2019
ASM800321v1	GCA_008003215.1	Wolbachia endosymbiont -W2.1	Annotation	10/05/2019

5) Pairwise genome alignment

Dot plot



Dgenies: <https://dgenies.toulouse.inra.fr>

Dot plot

In bioinformatics a dot plot is a graphical method that allows the comparison of two biological sequences and identify regions of close similarity between them. It is a type of recurrence plot.

More details of dot plot [here](#). Below, some examples of events which can be detected by dot plots.

Match

When two samples sequence are identical, it's a match.



Gap

Dot plots can be used to detect a gap between two samples: small sequence which exists only in one sample, between two matching regions.



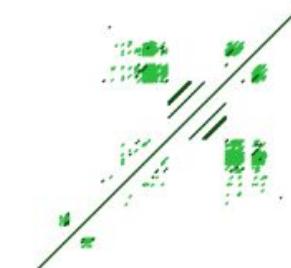
Inversion

Sequence which exists in the two samples but not in the same order.

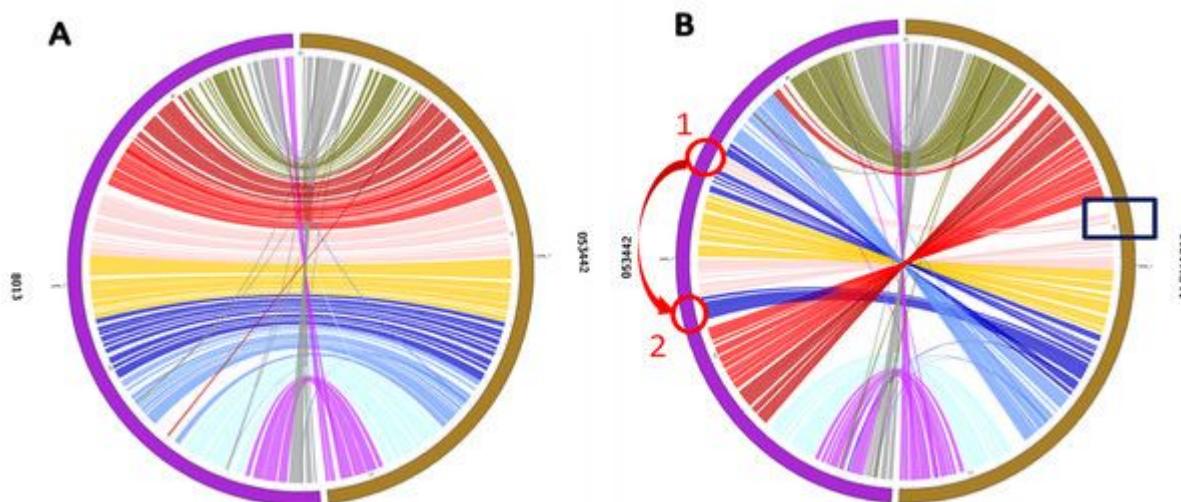


Repeats

Dot plot can be used to detect repeated regions: a sequence which is repeated several times in a sample.



Circos link

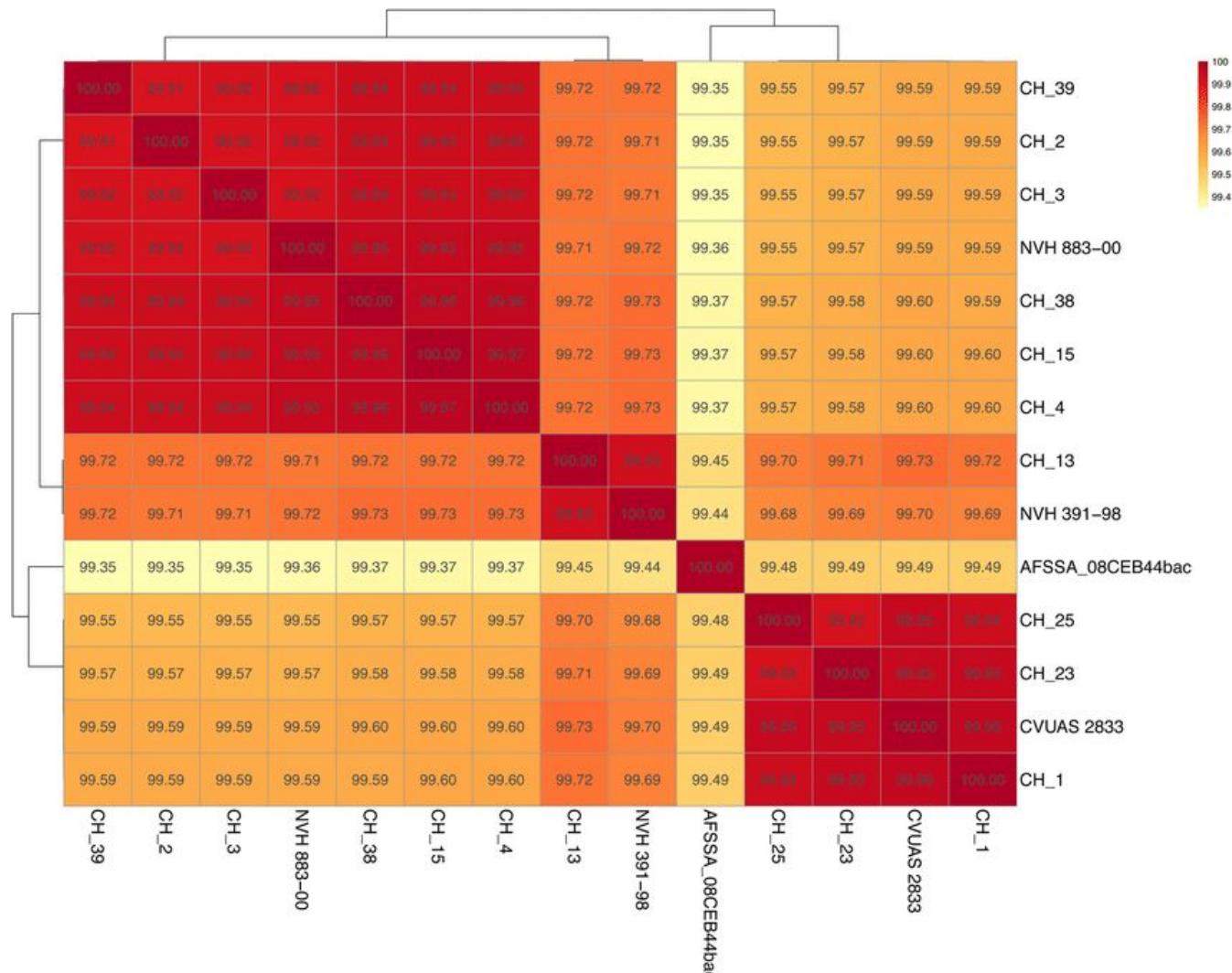


6) Pairwise Average Nucleotide Identity (ANI)

ANI: Average Nucleotide Identity

The average nucleotide identity (ANI) is a similarity index between a given pair of genomes that can be applicable to prokaryotic organisms independently of their G+C content, and a cutoff score of >95% indicates that they belong to the same species

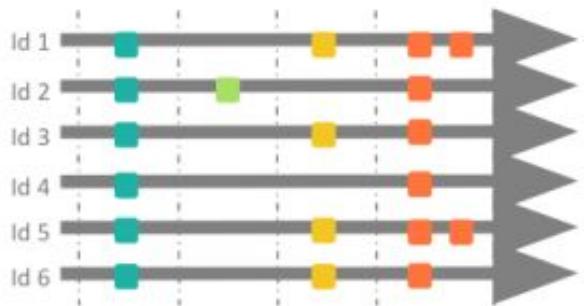
Program: FastANI



Heat map of the average nucleotide identity (ANI) for strains of the species *B. cytotoxicus* (Stevens et al., 2019)

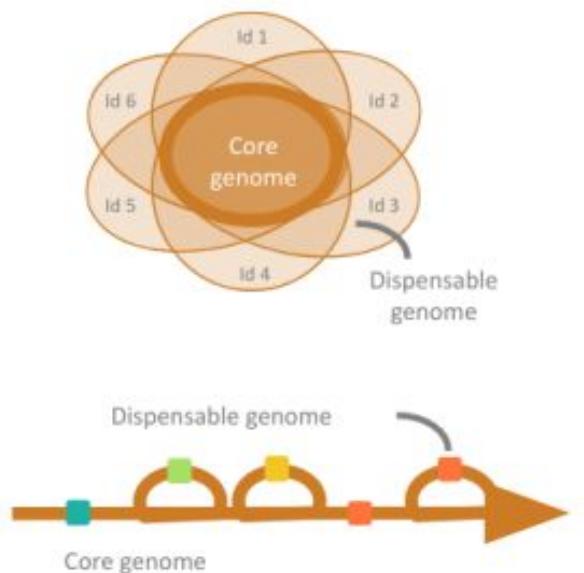
7) Pan-genome and Gene clustering

Pangenome concept



Pangenome

Collection of genes or sequences found in all individuals of a population (intra or inter species)



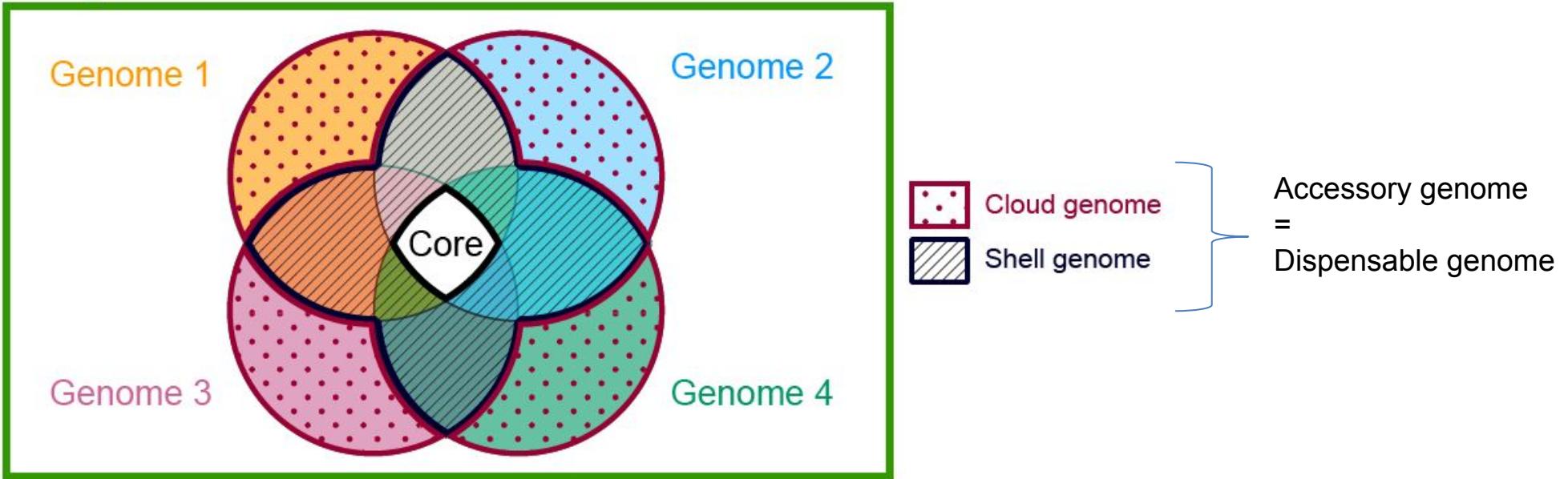
- ▶ **Core genome** : present in all individuals
- ▶ **Disposable genome** : absent from one or several individuals (also called variable, accessory,...)

Gene Clustering - how it works

- Assess the similarity of every gene to every other gene
 - e.g., using BLAST
- Use that similarity to join pairs of genes
 - e.g., using Reciprocal Best Hits
- Connect the gene pairs into larger clusters
 - e.g., using Reciprocal Best Hits or Markov clustering

=> Programs: OrthoMCL, Roary, PGAP...

Pangenome



Le pangénome ouvert, fermé, le ratio C/P

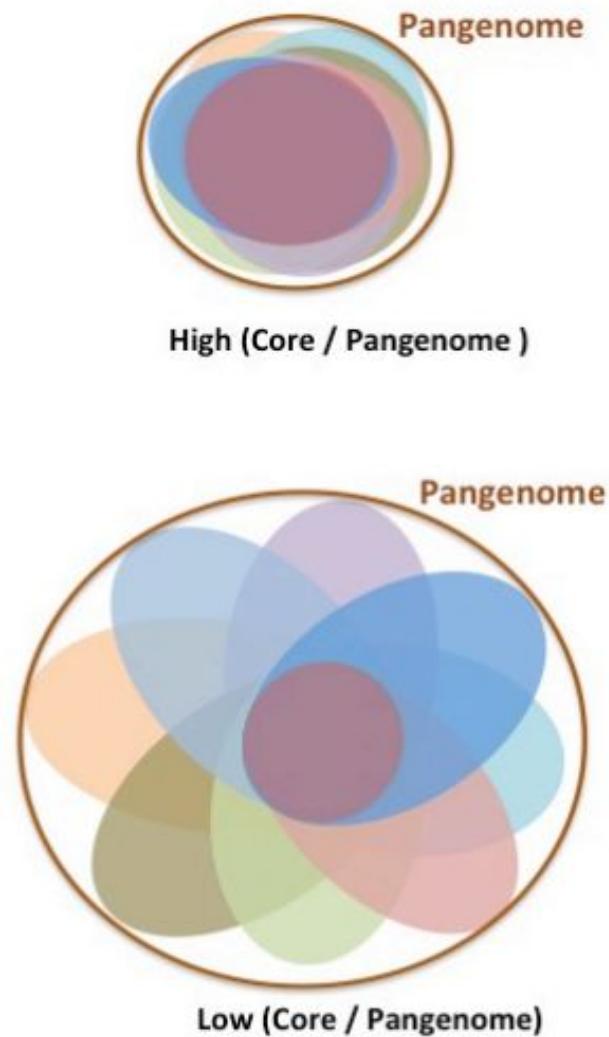
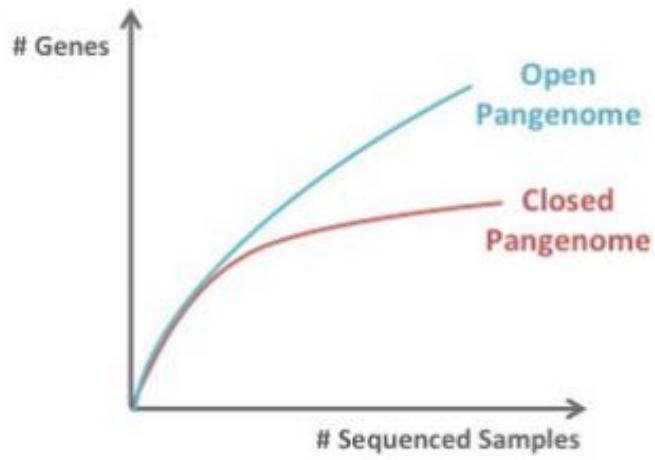


Table 1. Popular software for evolutionary pangenomics

Name	Authors	Reference
Panseq	Laing et al. (2010)	[12]
PanCGHweb	Bayjanov et al. (2010)	[13]
CAMBer	Wozniak et al. (2011)	[14]
PGAT	Brittnacher et al. (2011)	[15]
PGAP	Zhao et al. (2012)	[16]
GET_HOMOLOGUES	Contreras-Moreira and Vinuesa (2013)	[17]
GET_HOMOLOGUES-EST	Contreras-Moreira et al. (2017)	[18]
PanTools	Sheikhzadeh et al. (2016)	[19]
EDGAR 2.0	Blom et al. (2016)	[20]
PanX	Ding et al. (2018)	[21]
Micropan	Snipen and Liland (2015)	[22]
FindMyFriends	Pedersen (2015)	[23]
Piggy	Thorpe et al. (2018)	[24]
PanViz	Pedersen et al. (2017)	[25]

Method	Software	Input	Graph output	Pan-genome	Sequence homology	Paralogue identification
Roary	Conda package (v3.13.0)	GFF3	DOT	Directed graph	BLAST	Synteny
Ptolemy	Java executable (v1.0)	FASTA+GFF	GFA	Directed graph	minimap2	Graph-based
PPanGGoLin	Conda package (v1.0.13)	GBK or FASTA	GEXF	Undirected graph	MMseq2	Synteny
PIRATE	Conda package (v1.0.3)	GFF3	GFA	Directed graph	BLAST (/DIAMOND)	Synteny
Panaroo	Conda package (v1.1.2)	GFF3	GML	Directed graph	CD-HIT	Synteny

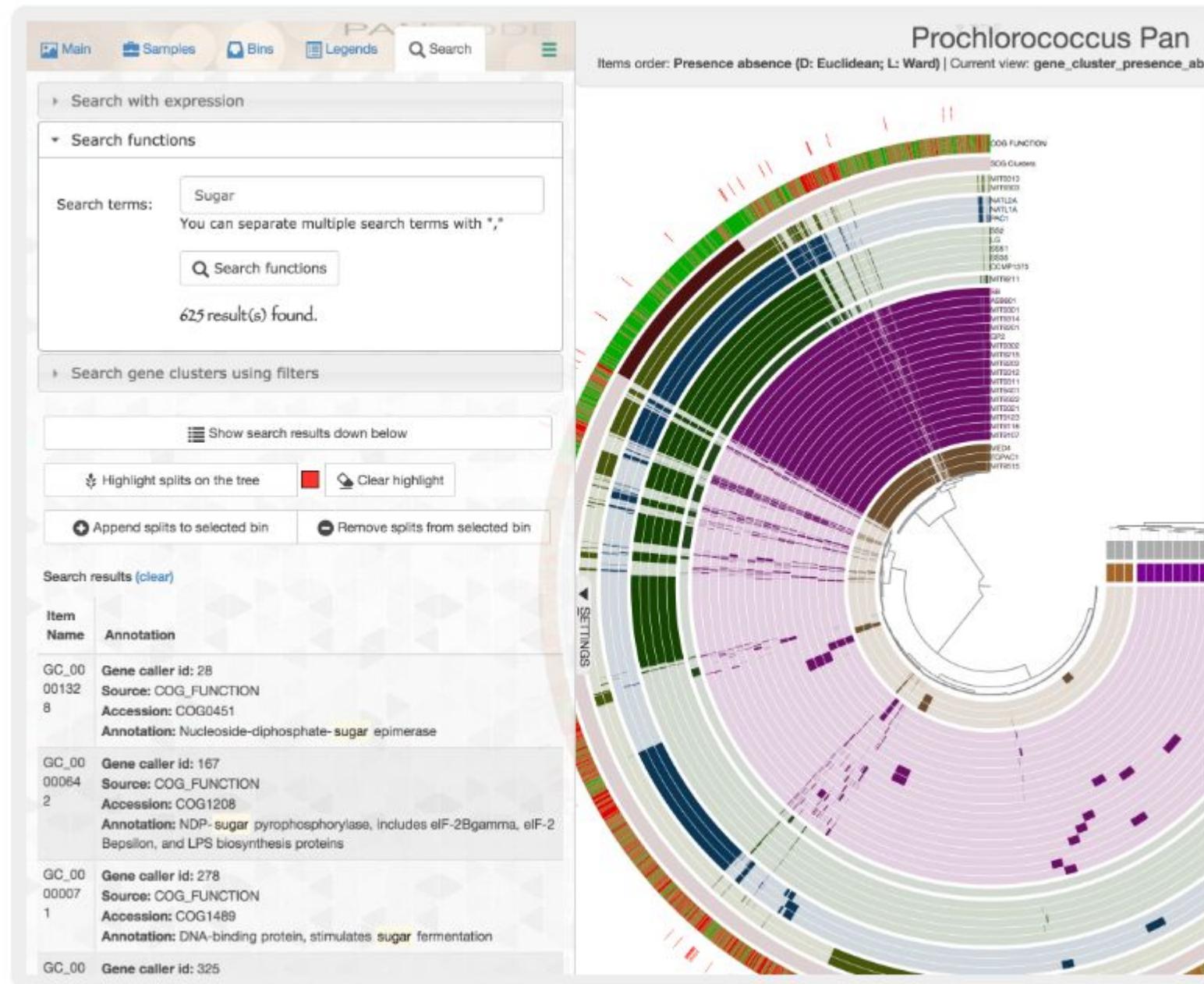
MICROBIAL GENOMICS

Volume 7, Issue 11

Research Article | Open Access

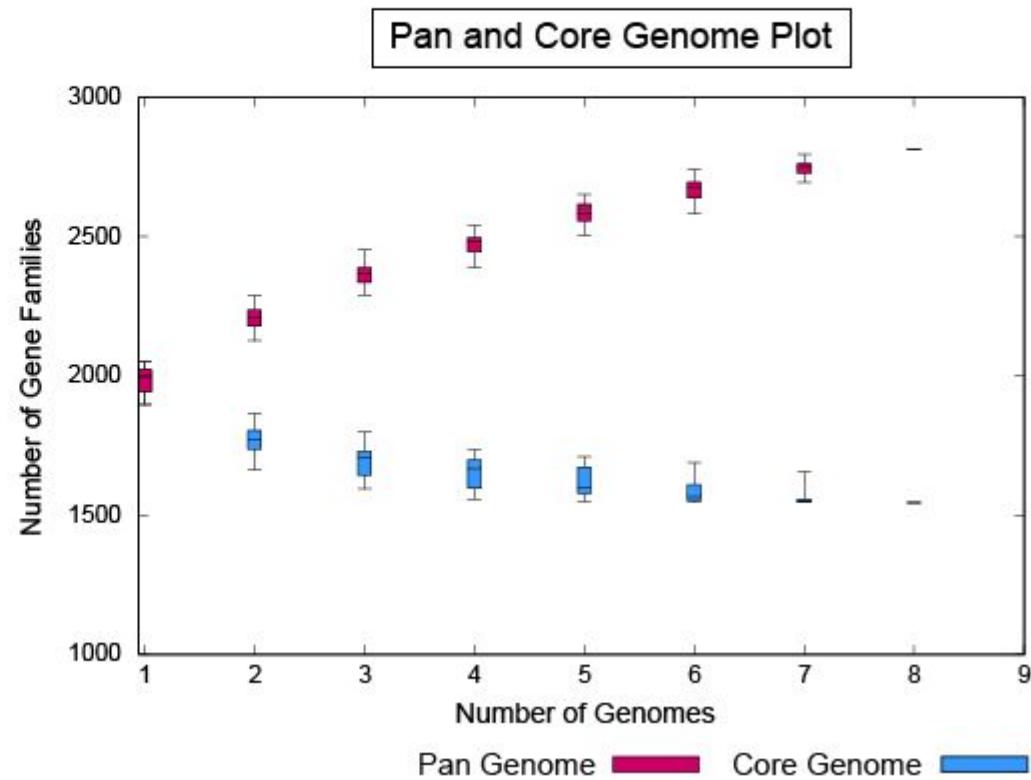
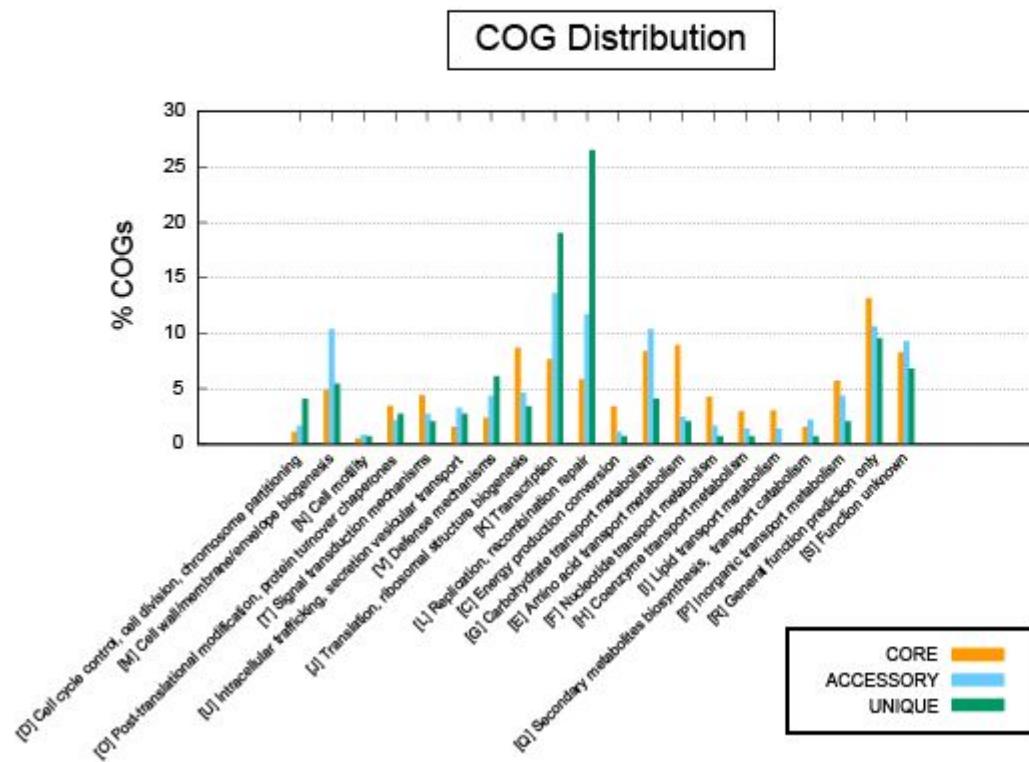
A comparative study of pan-genome methods for microbial organisms: *Acinetobacter baumannii* pan-genome reveals structural variation in antimicrobial resistance-carrying plasmids 

Aysun Urhan¹ , Thomas Aebel^{1,2} 

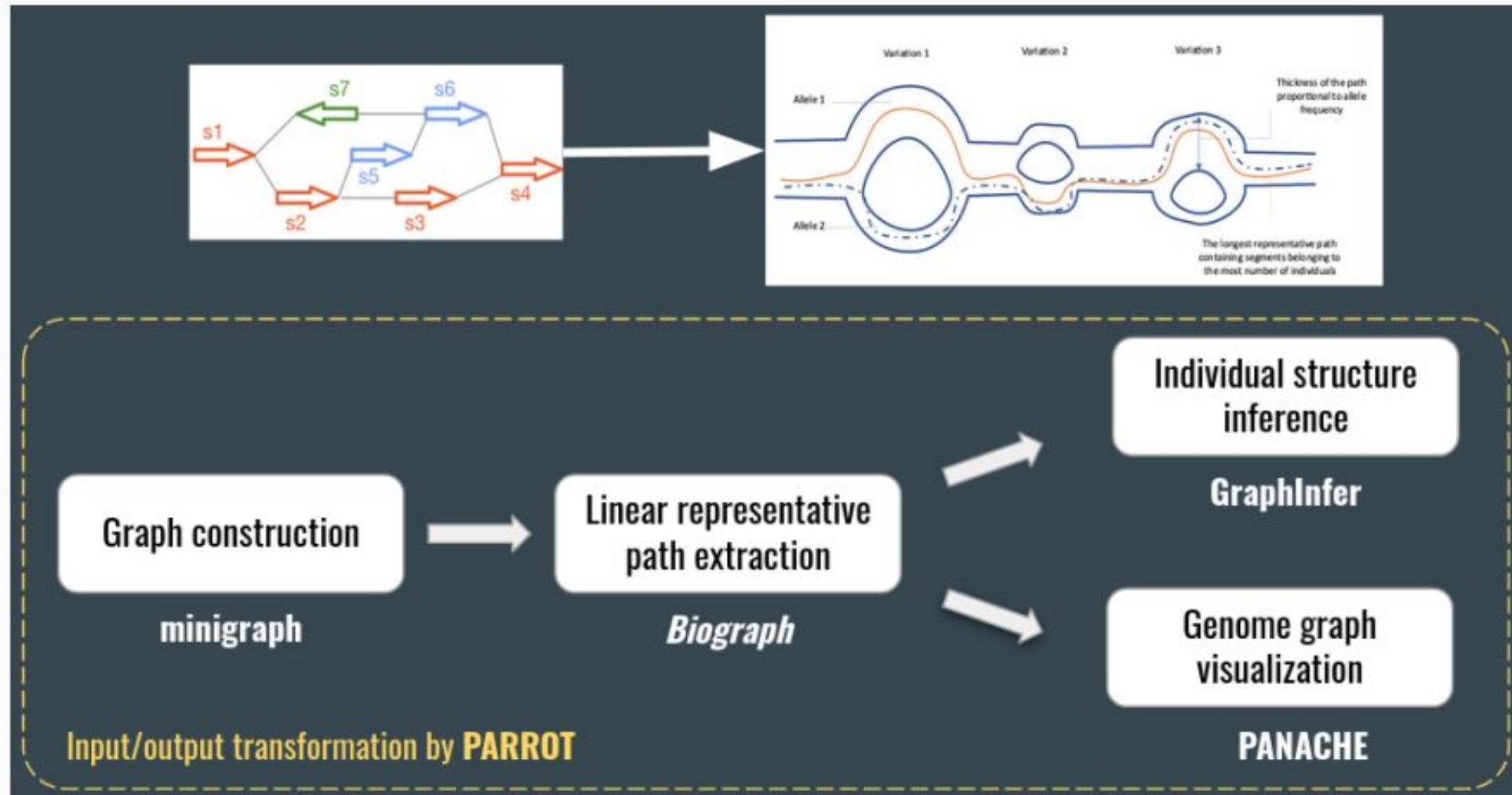


BPGA (Bacterial Pan Genome Analysis tool)

Streptococcus agalactiae

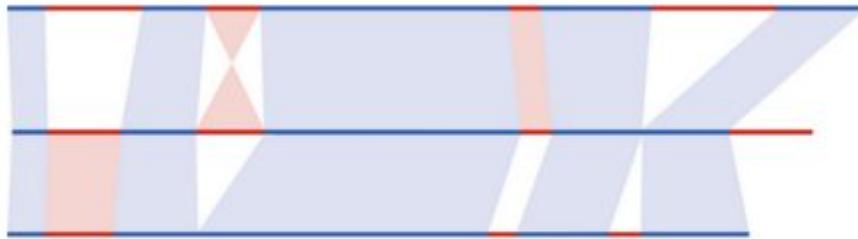


Comment manipuler le graphe pour les biologistes ?

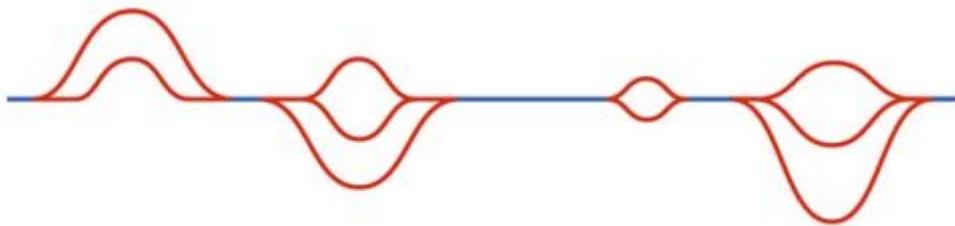


Concept du graphe de génome

Alignment of de novo assembled genomes



Pan-genome graph

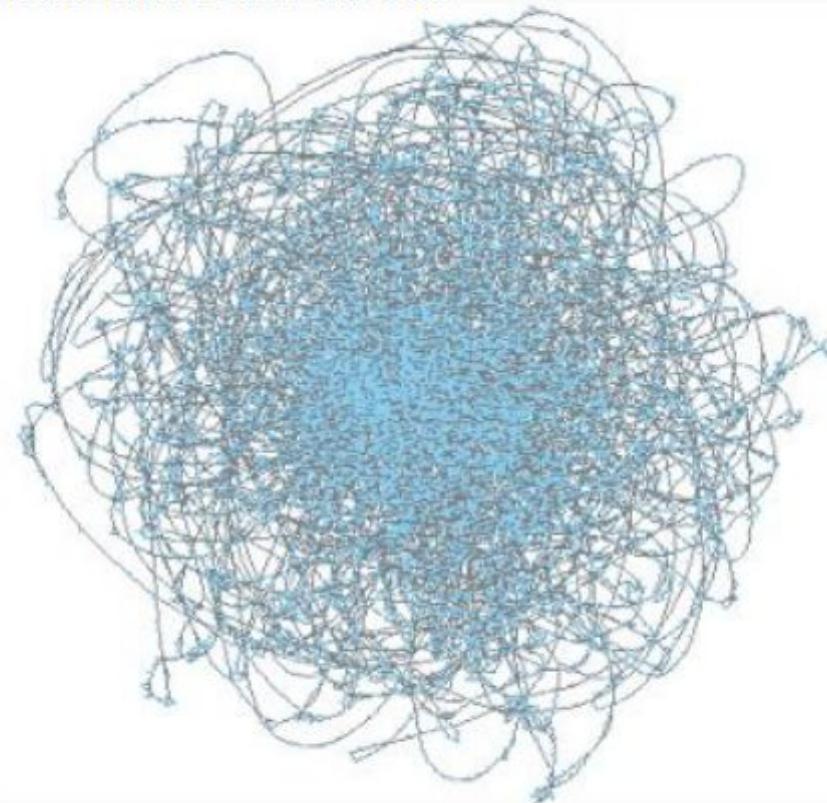


■ Dispensable genome

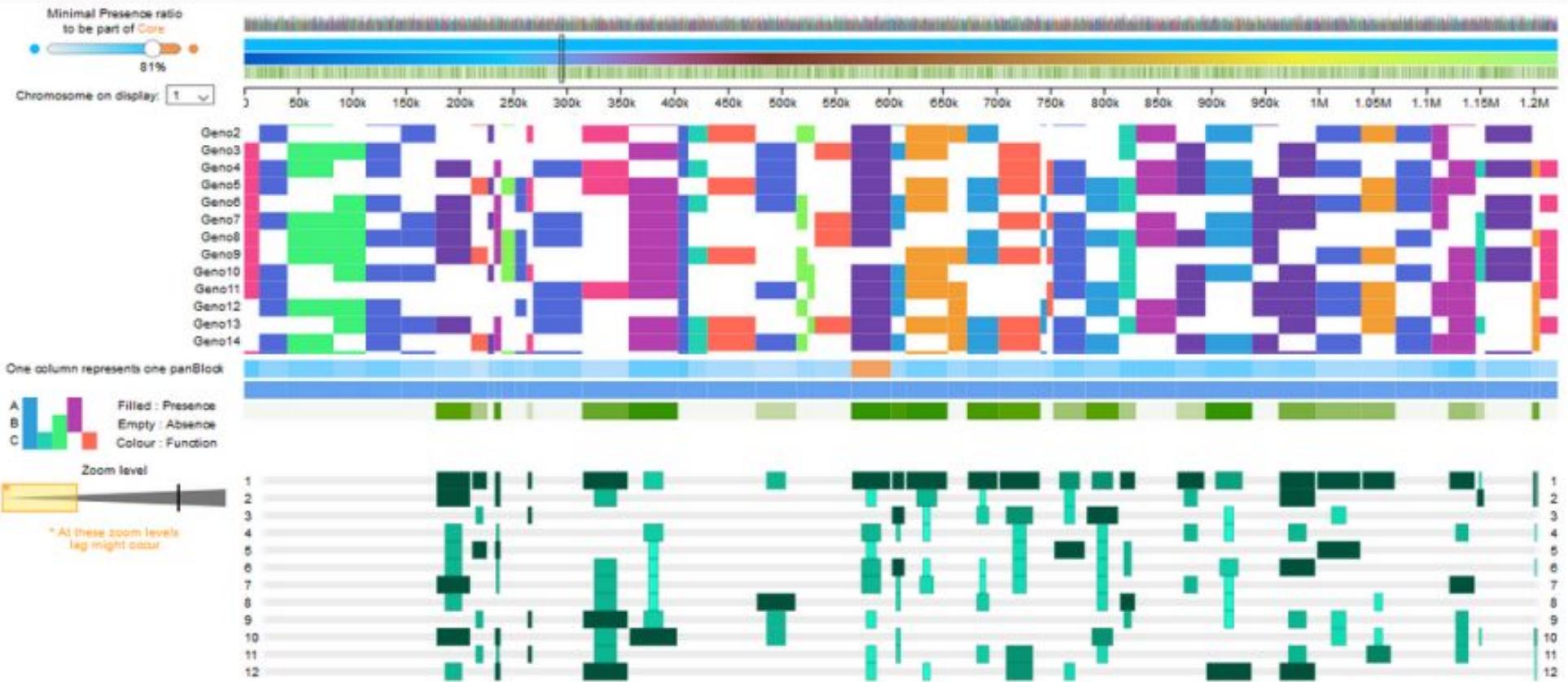
■ Core genome

Bayer et al., 2020

The HairBall effect



Un exemple linéaire, Panache



Durant, 2020-2021

8) Pan-GWAS

Pan-GWAS

Pan-GWAS of *Streptococcus agalactiae* Highlights Lineage-Specific Genes Associated with Virulence and Niche Adaptation

Authors: Andrea Gori , Odile B. Harrison, Ethwako Mlia, Yo Nishihara, Jia Mun Chan, Jacqueline Msefula, Macpherson Mallewa, [SHOW ALL \(13\)](#)

AUTHORS | Robert S. Heyderman | [AUTHORS INFO & AFFILIATIONS](#)

DOI: <https://doi.org/10.1128/mBio.00728-20>

 Check for updates

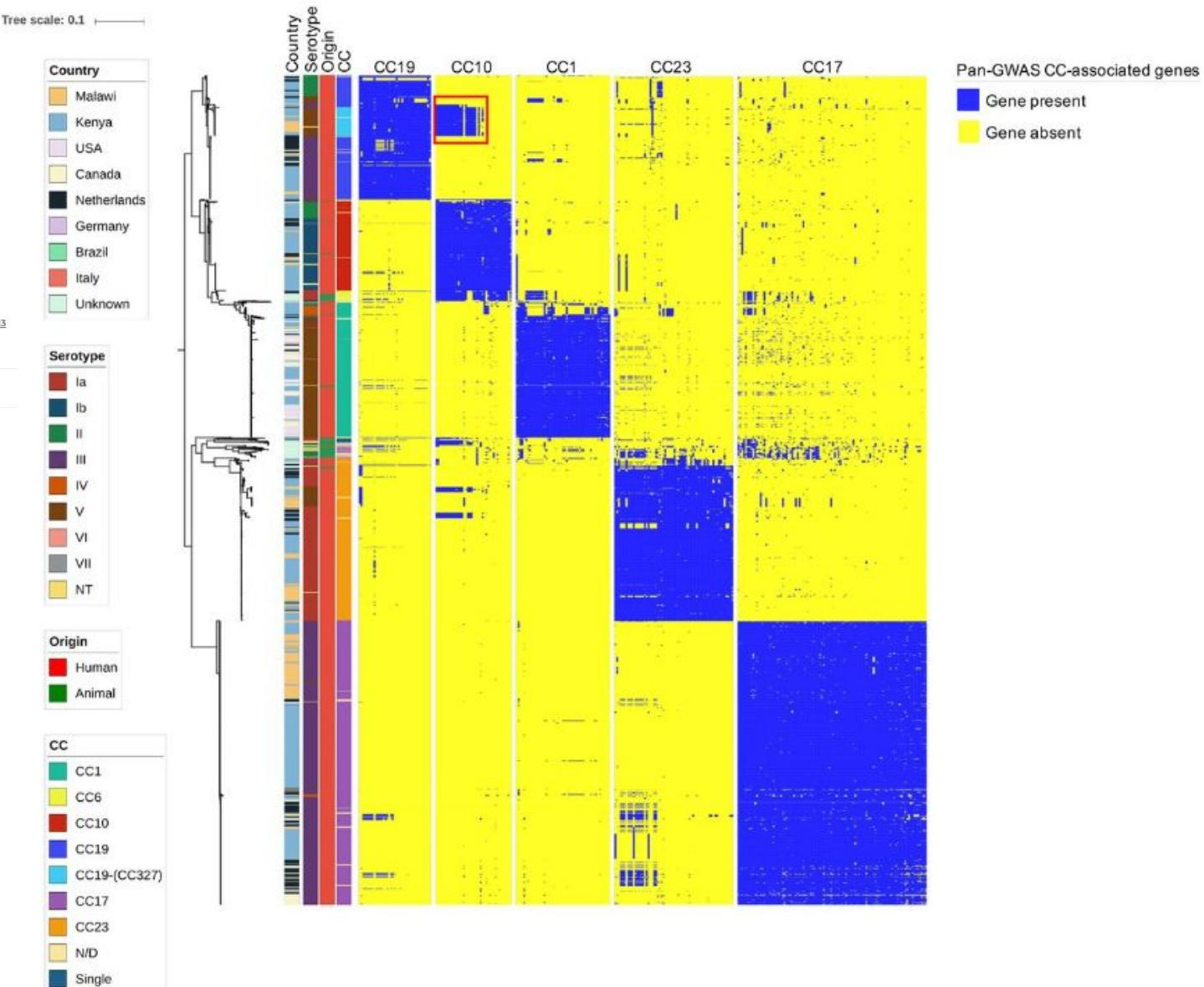


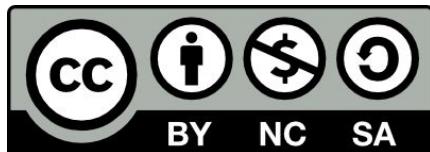
FIG 2 Core genome-based population structure of GBS. The phylogenetic tree is annotated with 4 colored strips representing the clonal complex, the country of isolation, the origin, and the serotype of each strain. The three binary heatmaps represent the presence (blue) or absence (yellow) of the genes identified by the pan-GWAS pipeline. The tree is rooted at midpoint. The reference strain used in this analysis was COH1, reference HG939456. The red square in the CC10 heatmap highlights the cluster of CC10-associated genes found in CC19 clones. Trees built with different reference strains are shown in Fig. S1 in the supplemental material and show analogous topology.

Odds ratios

Un *odds ratio* :

- < 1 signifie que l'événement est moins fréquent dans le groupe A que dans le groupe B ;
- = 1 signifie que l'événement est aussi fréquent dans les deux groupes ;
- > 1 signifie que l'événement est plus fréquent dans le groupe A que dans le groupe B.

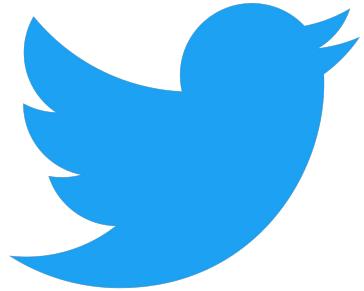
Merci pour votre attention !



Le matériel pédagogique utilisé pour ces enseignements est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions (BY-NC-SA) 4.0 International:
<http://creativecommons.org/licenses/by-nc-sa/4.0/>



SUIVEZ NOUS SUR TWITTER !



South Green : [@green_bioinfo](#)



i-Trop : [@ItropBioinfo](#)



N'oubliez pas de nous citer !

Comment citer les clusters?

"The authors acknowledge the IRD i-Trop HPC at IRD Montpellier for providing HPC resources that have contributed to the research results reported within this paper. URL: <http://bioinfo.ird.fr/> "

"The authors acknowledge the CIRAD UMR-AGAP HPC (South Green Platform) at CIRAD montpellier for providing HPC resources that have contributed to the research results reported within this paper. URL:
<http://www.southgreen.fr>"