

Pangénome et végétaux

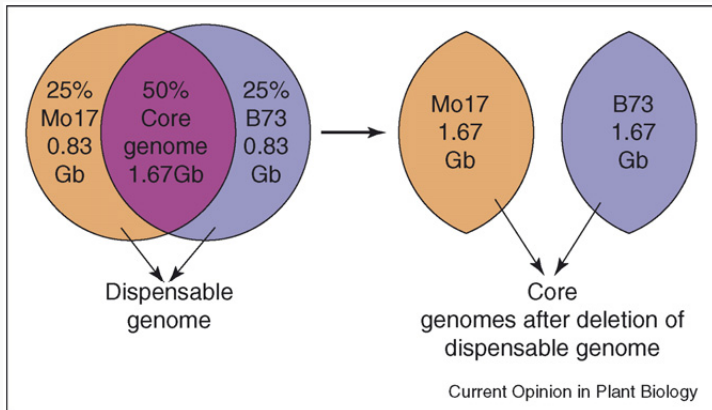
La vraie mesure de la diversité

Francois Sabot

13 Avril 2022

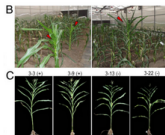
UMR DIADE, UM, IRD, CIRAD

Introduction



Morgante et al, 2007

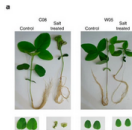
Le Pangénome, la vraie mesure de la diversité



From Yang et al., 2013



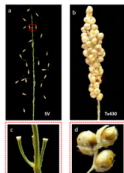
From Li et al. 2012



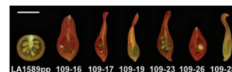
From Qi et al. 2014



From Yang et al., 2014



From Lin et al. 2012



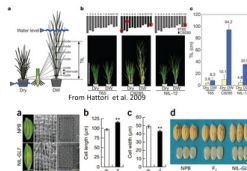
From Xiao et al. 2008



From Xu et al. 2016



From Gamuyao et al. 2012



From Hattori et al. 2009

From Wang et al. 2015



From Bai et al. 2017



- En 2020, 45 études (Monat et Sabot, 2020)

Pangenomics in Crop Plants



Cécile Monat and François Sabot

Abstract Identifying diversity at many scales is the keystone to understand how genomes evolved and plants are so adaptable, even if they are under extreme constraint. Understanding the processes that create diversity and how the genome manages this variability is of high importance to be able to face today's challenges for breeding and genetic resource conservation of crop plants. For about past 15 years, we ushered into the era of pangenomics and started to learn what the true intra-species genomic diversity is and how it shapes the structure and the expression of the genome, especially in crop plants. The pangenome is the complete repertoire of sequences for a given population (often a species). It can be divided in compartments, the first one is the core genome which contains the sequences shared by the whole population. The second compartment is the dispensable genome which regroups the sequences shared by some individuals of the population but not all of them. As a sub-part of the dispensable genome, an individual-specific-genome regroups the sequences that are uniquely found in one individual of the population. By comparing individual gene and sequence contents within a population, pangenomics is a great tool to investigate how sequences evolve within a species and how genetic diversity is shaped within it. More and more studies are including relative wild species to have a larger overview of how genetic and phenotypic diversity was shaped through domestication and selection processes. In this chapter, we will present this concept of pangenomics and the associated methods and progress made, particularly in crop plants and future perspective.

C. Monat
DIADE, University of Montpellier, IRD, Montpellier, France
South Green Bioinformatics Platform, IRD, Bioversity, CIRAD, INRAE, Montpellier, France
Université Clermont Auvergne, INRAE, GDEC, Clermont-Ferrand, France
F. Sabot (✉)
DIADE, University of Montpellier, IRD, Montpellier, France
South Green Bioinformatics Platform, IRD, Bioversity, CIRAD, INRAE, Montpellier, France
e-mail: Francois.sabot@ird.fr

Om P. Rajora (ed.), *Population Genomics: Crop Plants*,
Population Genomics [Om P. Rajora (Editor-in-Chief)],
https://doi.org/10.1007/13836_2020_84, © Springer Nature Switzerland AG 2020

- En 2020, 45 études (Monat et Sabot, 2020)
- En 2022, plus de 100!

Pangenomics in Crop Plants



Cécile Monat and François Sabot

Abstract Identifying diversity at many scales is the keystone to understand how genomes evolved and plants are so adaptable, even if they are under extreme constraint. Understanding the processes that create diversity and how the genome manages this variability is of high importance to be able to face today's challenges for breeding and genetic resource conservation of crop plants. For about past 15 years, we ushered into the era of pangenomics and started to learn what the true intra-species genomic diversity is and how it shapes the structure and the expression of the genome, especially in crop plants. The pangenome is the complete repertoire of sequences for a given population (often a species). It can be divided in compartments, the first one is the core genome which contains the sequences shared by the whole population. The second compartment is the dispensable genome which regroups the sequences shared by some individuals of the population but not all of them. As a sub-part of the dispensable genome, an individual-specific-genome regroups the sequences that are uniquely found in one individual of the population. By comparing individual gene and sequence contents within a population, pangenomics is a great tool to investigate how sequences evolve within a species and how genetic diversity is shaped within it. More and more studies are including relative wild species to have a larger overview of how genetic and phenotypic diversity was shaped through domestication and selection processes. In this chapter, we will present this concept of pangenomics and the associated methods and progress made, particularly in crop plants and future perspective.

C. Monat
DIADE, University of Montpellier, IRD, Montpellier, France
South Green Bioinformatics Platform, IRD, Bioversity, CIRAD, INRAE, Montpellier, France
Université Clermont Auvergne, INRAE, GDEC, Clermont-Ferrand, France
F. Sabot (✉)
DIADE, University of Montpellier, IRD, Montpellier, France
South Green Bioinformatics Platform, IRD, Bioversity, CIRAD, INRAE, Montpellier, France
e-mail: Francois.sabot@ird.fr

Om P. Rajora (ed.), *Population Genomics: Crop Plants*,
Population Genomics [Om P. Rajora (Editor-in-Chief)],
https://doi.org/10.1007/13836_2020_84, © Springer Nature Switzerland AG 2020

- En 2020, 45 études (Monat et Sabot, 2020)
- En 2022, plus de 100!
- De 3 à 3 000 individus

Pangenomics in Crop Plants



Cécile Monat and François Sabot

Abstract Identifying diversity at many scales is the keystone to understand how genomes evolved and plants are so adaptable, even if they are under extreme constraint. Understanding the processes that create diversity and how the genome manages this variability is of high importance to be able to face today's challenges for breeding and genetic resource conservation of crop plants. For about past 15 years, we ushered into the era of pangenomics and started to learn what the true intra-species genomic diversity is and how it shapes the structure and the expression of the genome, especially in crop plants. The pangenome is the complete repertoire of sequences for a given population (often a species). It can be divided in compartments, the first one is the core genome which contains the sequences shared by the whole population. The second compartment is the dispensable genome which regroups the sequences shared by some individuals of the population but not all of them. As a sub-part of the dispensable genome, an individual-specific-genome regroups the sequences that are uniquely found in one individual of the population. By comparing individual gene and sequence contents within a population, pangenomics is a great tool to investigate how sequences evolve within a species and how genetic diversity is shaped within it. More and more studies are including relative wild species to have a larger overview of how genetic and phenotypic diversity was shaped through domestication and selection processes. In this chapter, we will present this concept of pangenomics and the associated methods and progress made, particularly in crop plants and future perspective.

C. Monat
DIADE, University of Montpellier, IRD, Montpellier, France
South Green Bioinformatics Platform, IRD, Bioversity, CIRAD, INRAE, Montpellier, France
Université Clermont Auvergne, INRAE, GDEC, Clermont-Ferrand, France
F. Sabot (✉)
DIADE, University of Montpellier, IRD, Montpellier, France
South Green Bioinformatics Platform, IRD, Bioversity, CIRAD, INRAE, Montpellier, France
e-mail: Francois.sabot@ird.fr

Om P. Rajora (ed.), *Population Genomics: Crop Plants*,
Population Genomics [Om P. Rajora (Editor-in-Chief)],
https://doi.org/10.1007/13836_2020_84, © Springer Nature Switzerland AG 2020

- En 2020, 45 études (Monat et Sabot, 2020)
- En 2022, plus de 100!
- De 3 à 3 000 individus
- Génomes, transcriptomes, lectures courtes, longues, assemblages, mapping...

Pangenomics in Crop Plants



Cécile Monat and François Sabot

Abstract Identifying diversity at many scales is the keystone to understand how genomes evolved and plants are so adaptable, even if they are under extreme constraint. Understanding the processes that create diversity and how the genome manages this variability is of high importance to be able to face today's challenges for breeding and genetic resource conservation of crop plants. For about past 15 years, we ushered into the era of pangenomics and started to learn what the true intra-species genomic diversity is and how it shapes the structure and the expression of the genome, especially in crop plants. The pangenome is the complete repertoire of sequences for a given population (often a species). It can be divided in compartments, the first one is the core genome which contains the sequences shared by the whole population. The second compartment is the dispensable genome which regroups the sequences shared by some individuals of the population but not all of them. As a sub-part of the dispensable genome, an individual-specific-genome regroups the sequences that are uniquely found in one individual of the population. By comparing individual gene and sequence contents within a population, pangenomics is a great tool to investigate how sequences evolve within a species and how genetic diversity is shaped within it. More and more studies are including relative wild species to have a larger overview of how genetic and phenotypic diversity was shaped through domestication and selection processes. In this chapter, we will present this concept of pangenomics and the associated methods and progress made, particularly in crop plants and future perspective.

C. Monat
DIADE, University of Montpellier, IRD, Montpellier, France
South Green Bioinformatics Platform, IRD, Bioversity, CIRAD, INRAE, Montpellier, France
Université Clermont Auvergne, INRAE, GDEC, Clermont-Ferrand, France
F. Sabot (✉)
DIADE, University of Montpellier, IRD, Montpellier, France
South Green Bioinformatics Platform, IRD, Bioversity, CIRAD, INRAE, Montpellier, France
e-mail: Francois.sabot@ird.fr

Om P. Rajora (ed.), *Population Genomics: Crop Plants*,
Population Genomics [Om P. Rajora (Editor-in-Chief)],
https://doi.org/10.1007/13836_2020_84, © Springer Nature Switzerland AG 2020

Assemble-then-map

Whole genome



Song et al.
2020



Gao et al.
2019

Draft assembly*



Schatz et al., 2014
Wang et al. 2018
Zhao et al. 2018

Graph



Liu et al.
2020



Gordon et al.
2017



Walkowiak et al.
2020

Map-then-assemble

Iterative approach



Golicz et al.
2016



Rijzaanj et al.
2021

Bulk



Yao et al.
2018



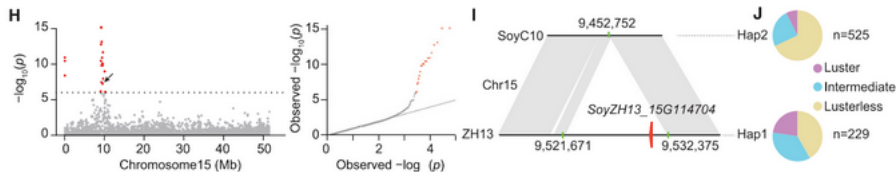
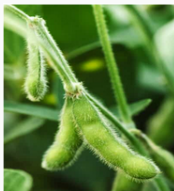
Montenegro et al.
2018

Parallel approach



Hubner et al.
2019

Quelques exemples icôniques



Liu et al, 2020

27 175 gènes absent de la référence (26 cultivars)



Montenegro et al, 2017

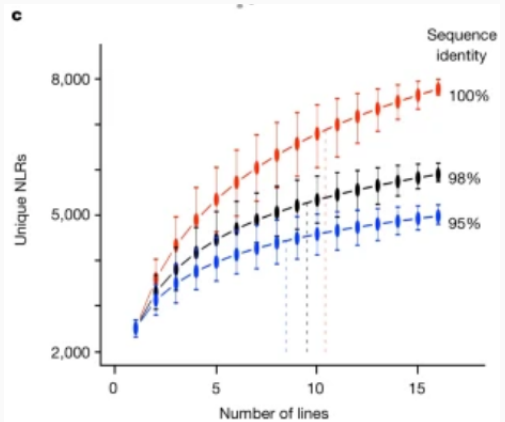
18 cultivars, 12 150 gènes absents de la référence

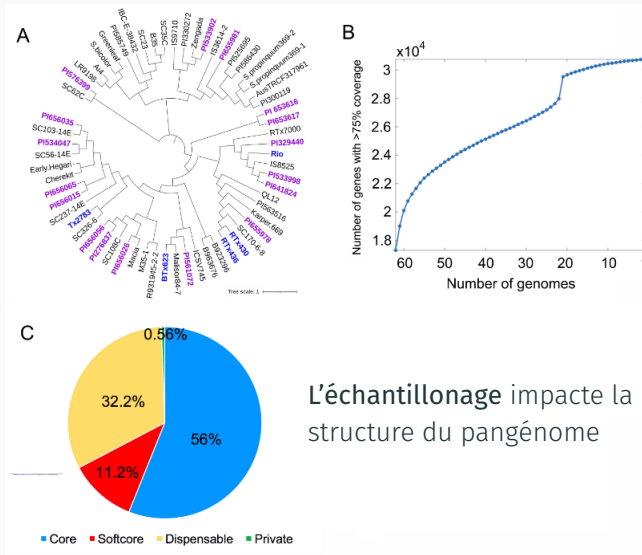


Montenegro et al, 2017

Limite méthodologique de la
définition du pangénome

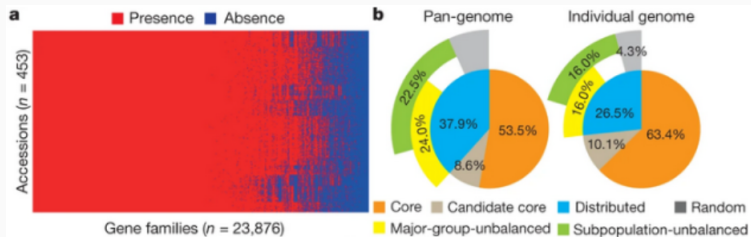
18 cultivars, 12 150 gènes absents de la référence





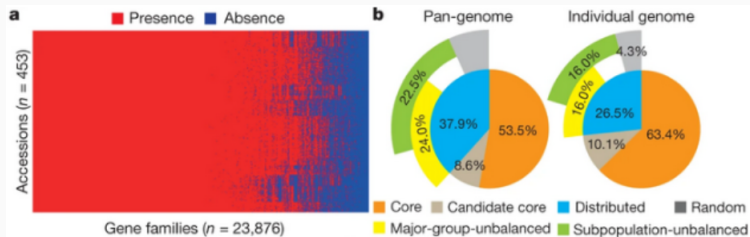
Wang et al, 2021

L'échantillonnage impacte la structure du pangénome



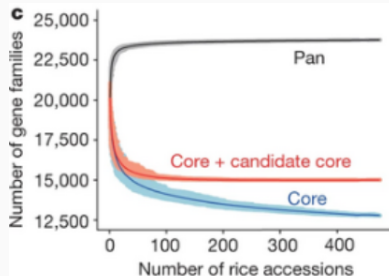
Wang et al, 2018

Le riz asiatique, histoire complexe



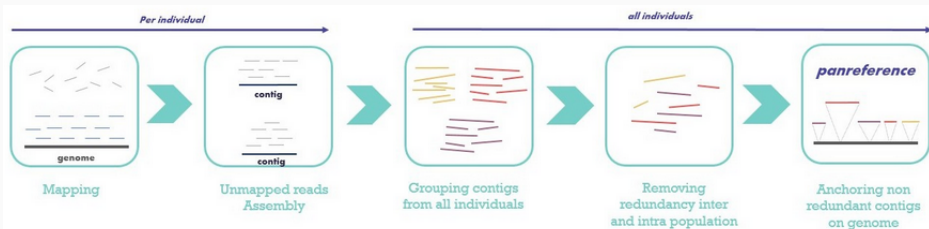
Wang et al, 2018

Le core diminue de plus en plus...



Wang et al, 2018

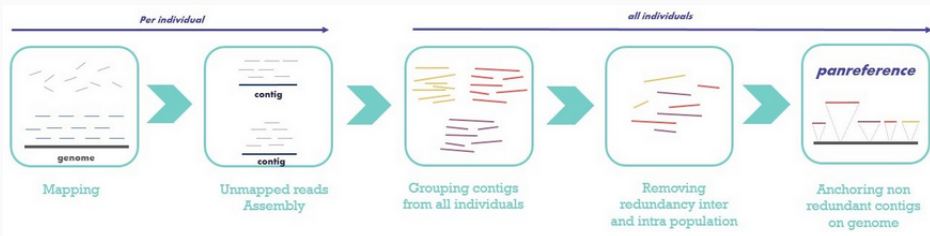
Qu'y a-t'il dans le dispensable du riz africain ?



Tranchant-Dubreuil et al, unpublished



Qu'y a-t'il dans le dispensable du riz africain ?

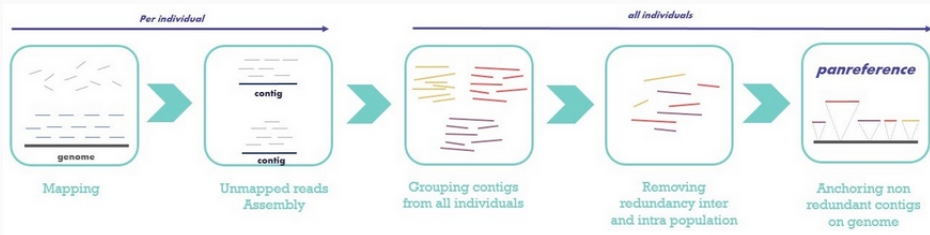


Tranchant-Dubreuil et al, unpublished

- 484 394 séquences (300-83kb)
- 515 Mb non redondant (350Mb/genomes)



Qu'y a-t'il dans le dispensable du riz africain ?

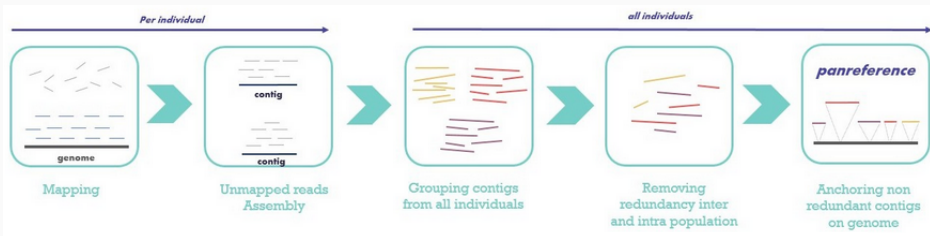


Tranchant-Dubreuil et al, unpublished

- 484 394 séquences (300-83kb)
- 515 Mb non redondant (350Mb/genomes)
- 67% d'éléments transposables



Qu'y a-t'il dans le dispensable du riz africain ?



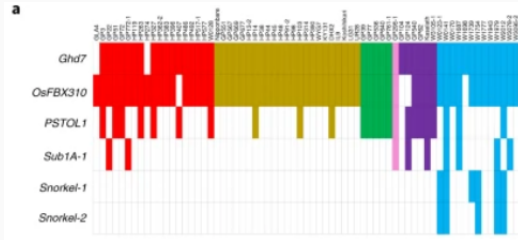
Tranchant-Dubreuil et al, unpublished

- 484 394 séquences (300-83kb)
- 515 Mb non redondant (350Mb/genomes)
- 67% d'éléments transposables
- 2 500 gènes, dont 2 106 sous sélection

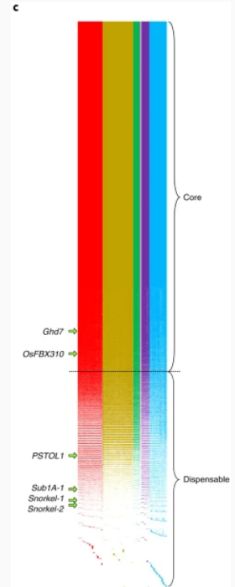


Impact de la domestication

Le riz, cultivé et sauvage

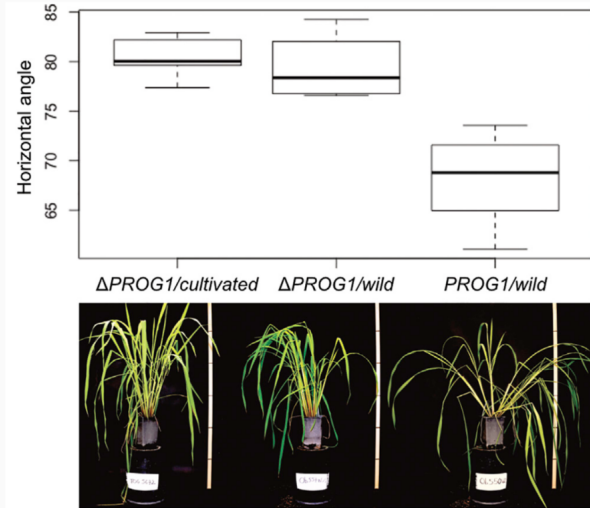


Zhao et al, 2018



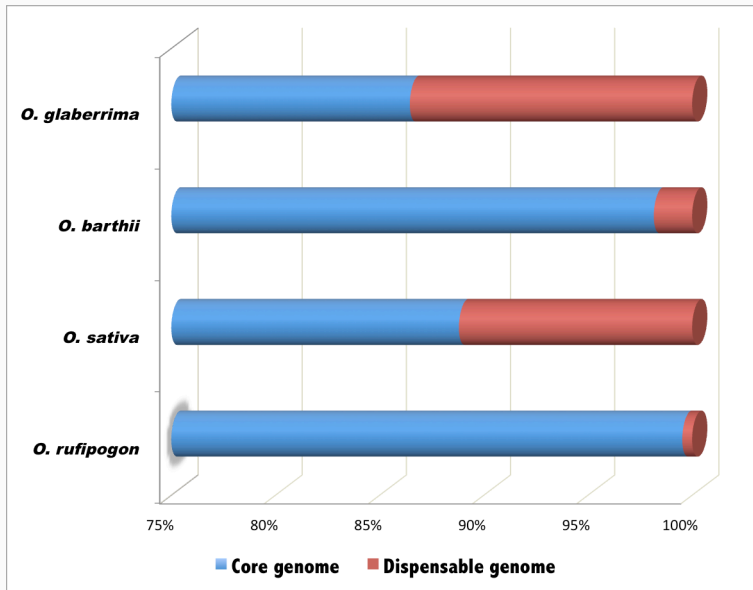


Fort impact des domestications chez le riz

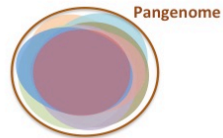
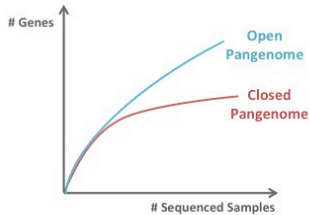


From Cubry et al, 2018

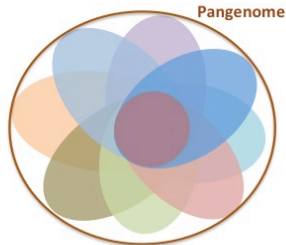
Fort impact des domestications chez le riz



Le pangénome ouvert, fermé, le ratio C/P



High (Core / Pangenome)

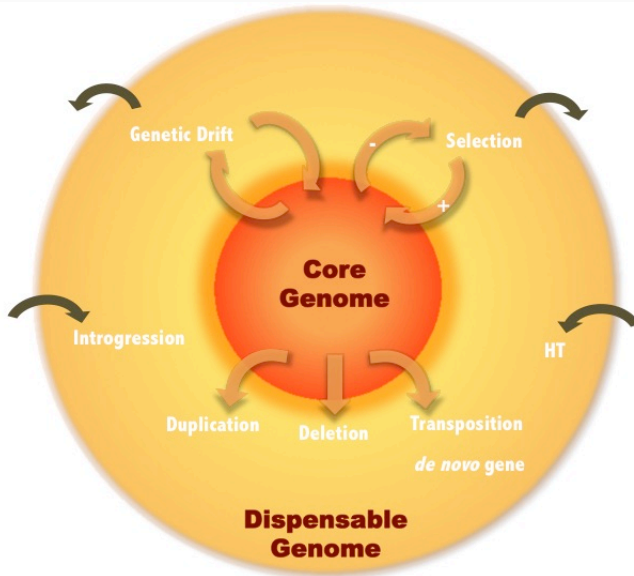


Low (Core / Pangenome)

Tranchant-Dubreuil, Rouard & Sabot, 2018



Monat & Sabot, unpublished



Représentation & Visualisation

Concept du graphe de génome

Alignment of de novo assembled genomes



Pan-genome graph



■ Dispensable genome ■ Core genome

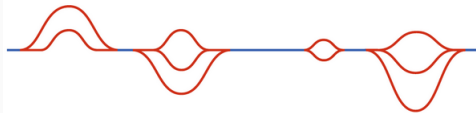
Bayer et al., 2020

Concept du graphe de génome

Alignment of de novo assembled genomes



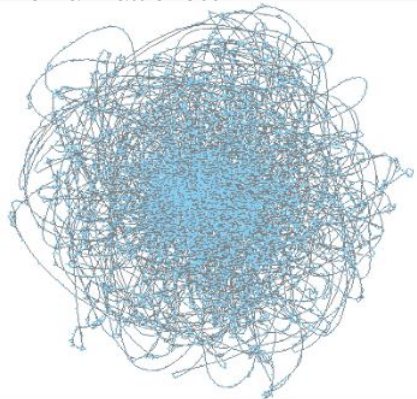
Pan-genome graph

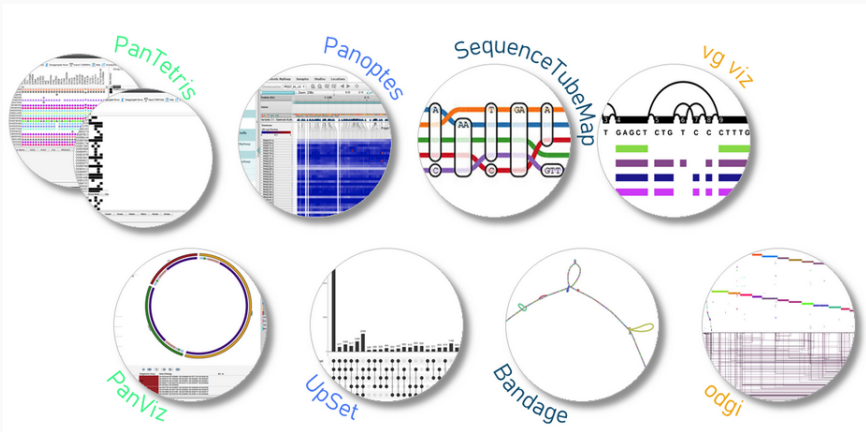


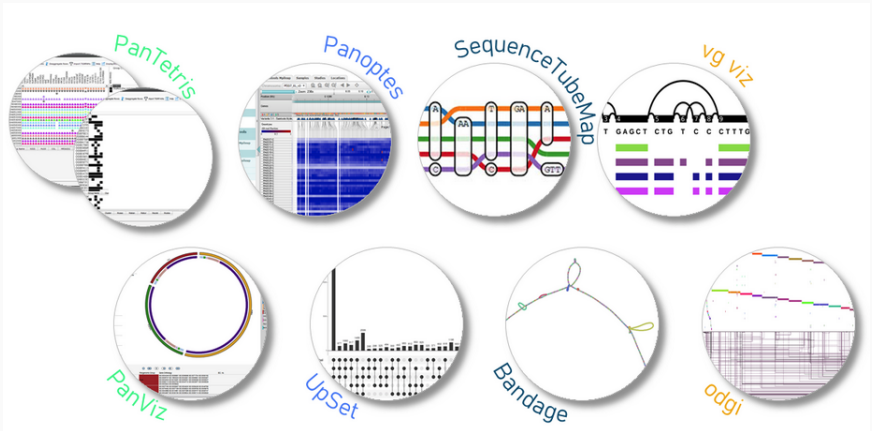
■ Dispensable genome ■ Core genome

Bayer et al., 2020

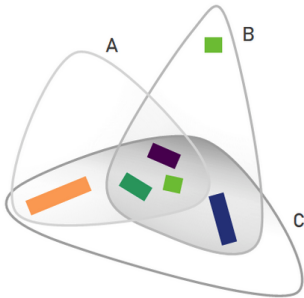
The HairBall effect



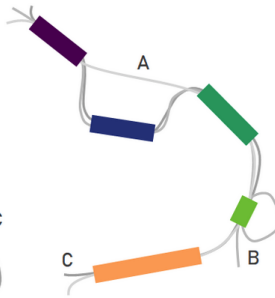




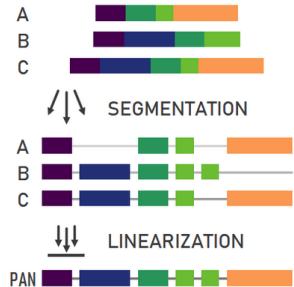
Problèmes de lisibilité, passage à l'échelle...
Besoin de nouveaux paradigmes de visualisation



Sets



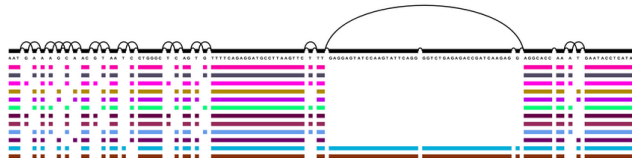
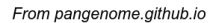
Graphs



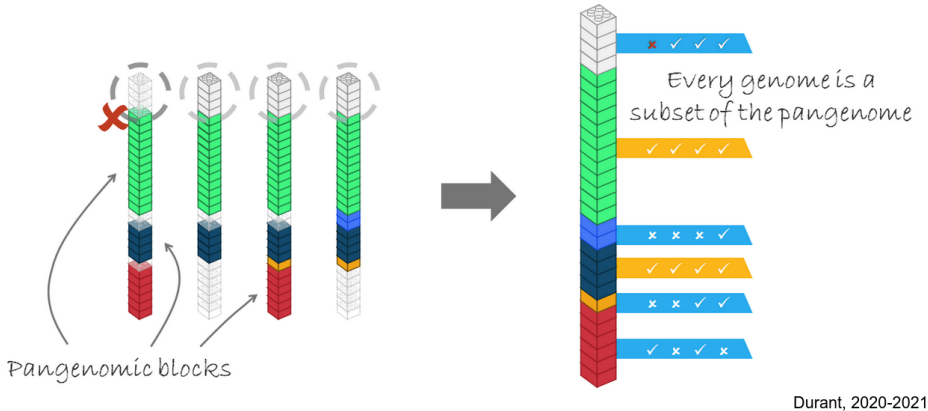
Linearity

Durant, 2020-2021

IRD
Institut de Recherche
pour le Développement
FRANCE



Tout génome est un *subset* du pangenome



Durant, 2020-2021

Un exemple linéaire, Panache



Durant, 2020-2021

Conclusions

- Une re-découverte de la diversité végétale

- Une re-découverte de la diversité végétale
- Pas mal de limitations techniques (en cours de résolution)

- Une re-découverte de la diversité végétale
- Pas mal de limitations techniques (en cours de résolution)
- La visualisation, *The Next Frontier*

- Une re-découverte de la diversité végétale
- Pas mal de limitations techniques (en cours de résolution)
- La visualisation, *The Next Frontier*
- Mais toujours la même idée:
 - Un seul génome est loin d'être suffisant

- Une re-découverte de la diversité végétale
- Pas mal de limitations techniques (en cours de résolution)
- La visualisation, *The Next Frontier*
- Mais toujours la même idée:
 - Un seul génome est loin d'être suffisant
 - Les allèles ne sont pas tout

- Une re-découverte de la diversité végétale
- Pas mal de limitations techniques (en cours de résolution)
- La visualisation, *The Next Frontier*
- Mais toujours la même idée:
 - Un seul génome est loin d'être suffisant
 - Les allèles ne sont pas tout
- **Le Pangénome est la clef pour une agriculture résiliente**

- Thi-Minh Nguyet Dang (graphe, linéarisation)
- Tuan Do (graphe)
- Eloi Durant (maître ès visualisation)
- Cécile Monat (méthodes Core)
- Christine Tranchant-Dubreuil (méthodes Dispensable)
- J Orjuela, M Rouard, M Conte, L Albar, C Muller...



Merci pour votre attention