Séquençage mate-pair

Assemblage et variabilité de structure

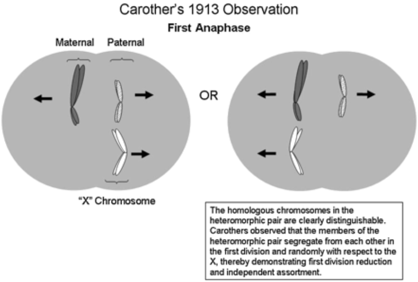
Sylvain Foissac[1], Diane Esquerré[2], Katia Fève[1], Juliette Riquet[1] et Thomas Faraut[1]

[1]Laboratoire de Génétique Cellulaire
[2]Plateforme Get-PlaGe
INRA, Toulouse

Toulouse, 30 septembre 2013

Théorie chromosomique de l'hérédité

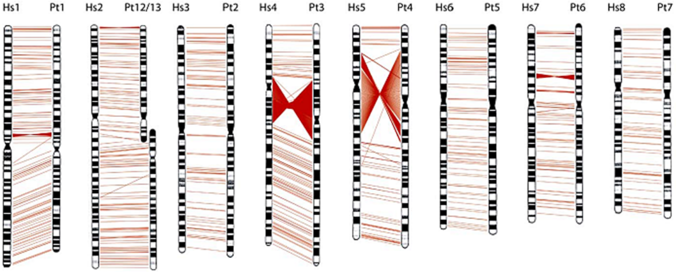
Estella Eleanor Carothers (1883-1957)



The homologous chromosomes in the heteromorphic pair are clearly distinguishable. Carothers observed that the members of the heteromorphic pair segregate from each other in the first division and randomly with respect to the X, thereby demonstrating first division reduction and independent assortment.

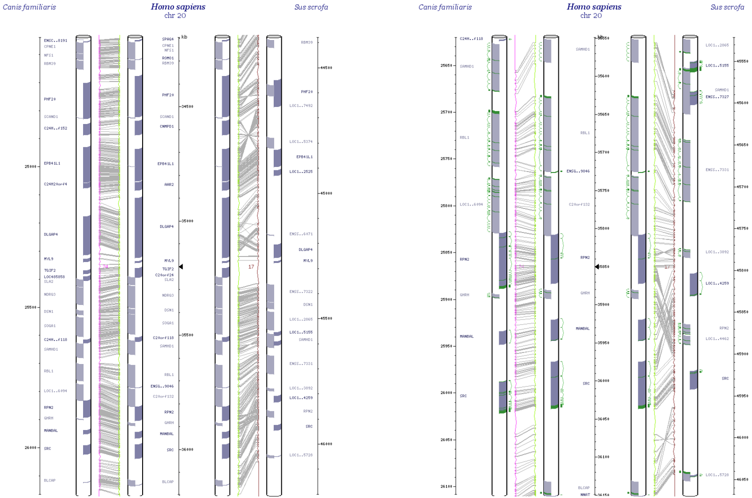
Variabilité de structure

> *"[...] we're missing 50% of the variants out there because of the limitation of our methods and our technology."*
>
> *E. Eichler in Baker* et al. *Nat Meth (2012)*



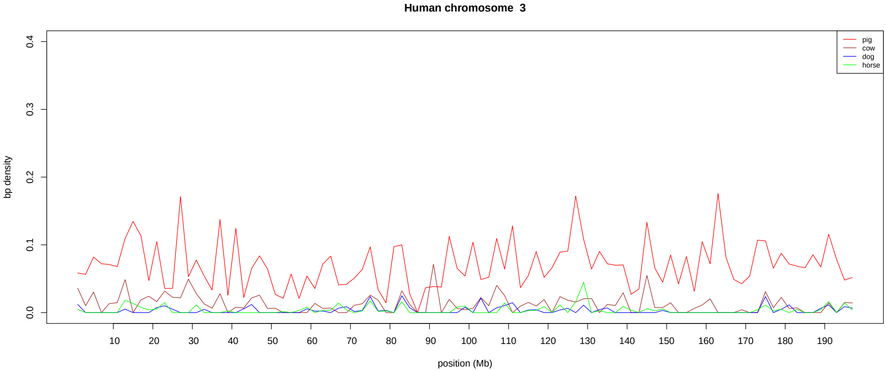
Feuk *et al.* Plos Genet (2005)

Assemblage porcin Sscrofa10.2



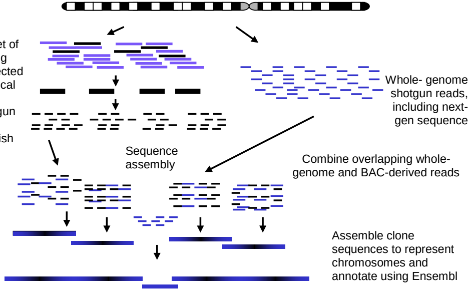
http://narcisse.toulouse.inra.fr

Assemblage porcin Sscrofa10.2



Assemblage porcin Sscrofa10.2

Le génome porcin a été séquencé en utilisant une approche *shotgun* hiérarchique :



Minimal set of overlapping BACs selected from physical

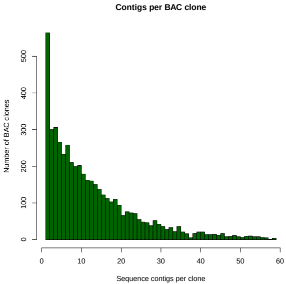
BAC shotgun reads
4x - prefinish

Whole- genome shotgun reads, including next-gen sequence

Sequence assembly

Combine overlapping whole-genome and BAC-derived reads

Assemble clone sequences to represent chromosomes and annotate using Ensembl

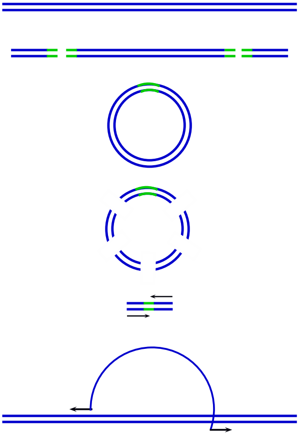


Archiblad *et al.* BMC Genomics (2010)

Projet PlastiPig

- Détecter les variants de structure chez le cochon à l'aide de séquençage mate-pair
- Pour chaque race (Duroc, Large White, Pietrain, Landrace, Meishan, ...), une librairie mate-pair est construite à partir d'un pool de 5 animaux
- Exploiter ces données pour améliorer l'assemblage porcin

Librairie mate-pair (Nextera)



tagmenter (6kb)
(adaptateurs biotynilés)

↓

circulariser

↓

fragmenter (500-600bp)

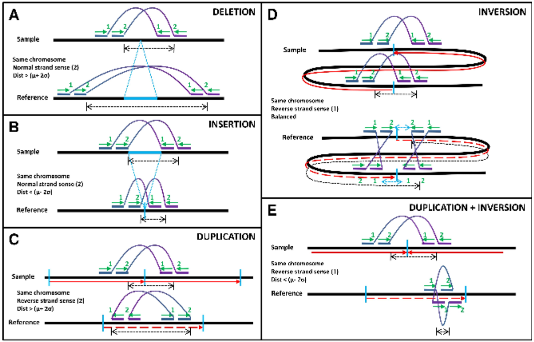
↓

sequencer

↓

aligner

Mate-pair et variants de structure



Zeitouni *et al.* Bioinformatics (2010)

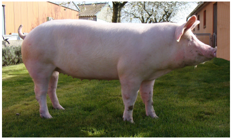
Données

Piétrain

153,631,248 lectures

Duroc

237,351,966 lectures



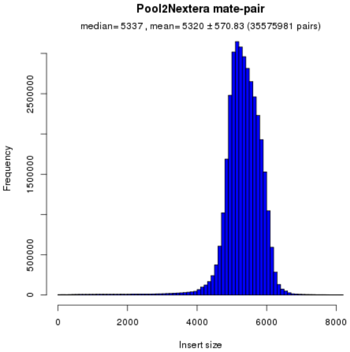
Large White

179,873,726 lectures

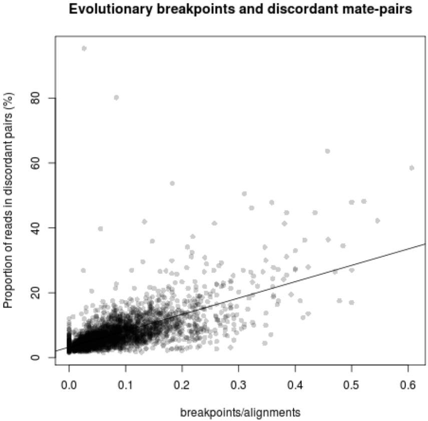
Données

La librairie Piétrain

Reads		
total	153,631,248	100%
mapped	119,273,286	77.6%
duplicates	12,439,227	8.0%
mapped nodup	89,797,907	58.4%
Mapped pairs		
total unique	44,898,953	100%
trans chrom	2,175,820	4.8%
cis	42,723,133	95.1%
same sense	711,677	1.5%
convergent	41,766,481	93.0%
proper pairs	41,436,485	92.2%
divergent	244,975	0.5%
paired-ends	19,680	0.0004%

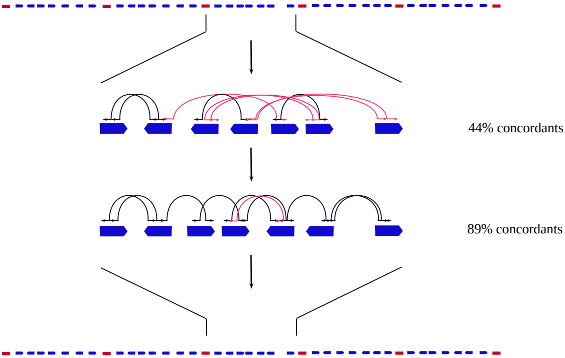


Mate-pair et points de cassure



Algorithme

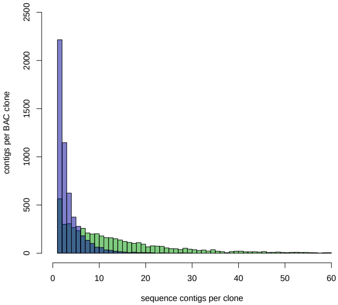
```
for each chromosome do
    for each clone do
        extract inter-contig discordant mate-pairs
        reassemble them using opera
    end for
    concatenate clones back into chromosomes
end for
```

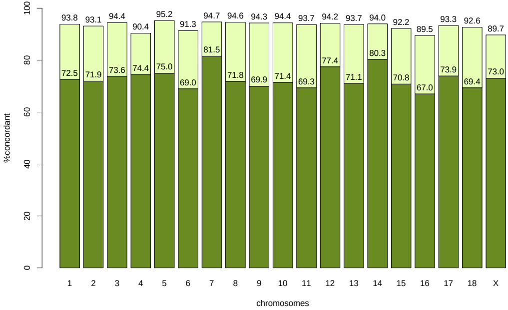


Gao, Sung et Nagarajang. Opera. J Comp Biol(2011)

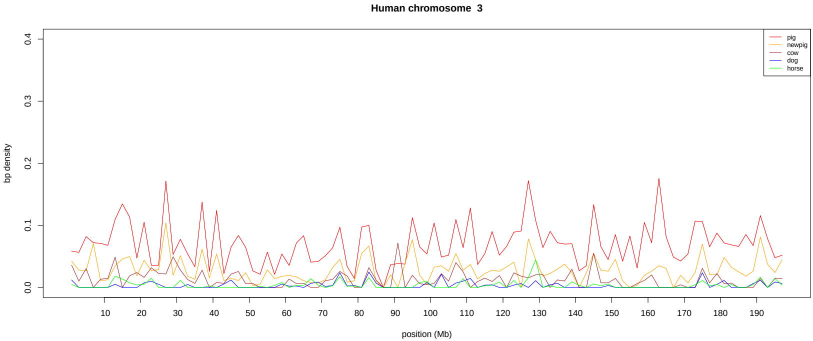
Résultats

De 73,372 contigs à 15,522 contigs

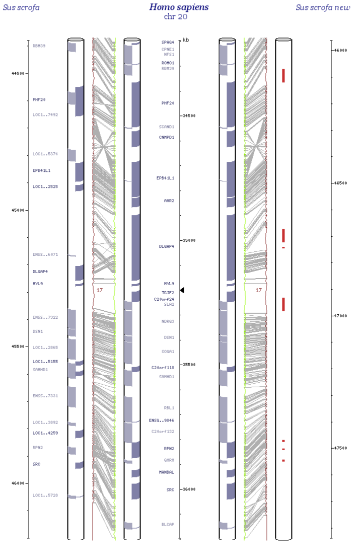


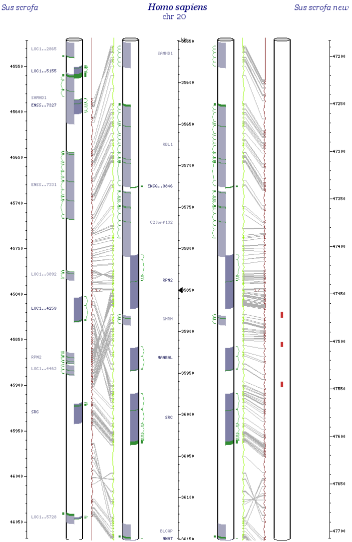


Résultats


Human chromosome 3
bp density
0.0
0.1
0.2
0.3
0.4
position (Mb)
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
pig
newpig
cow
dog
horse

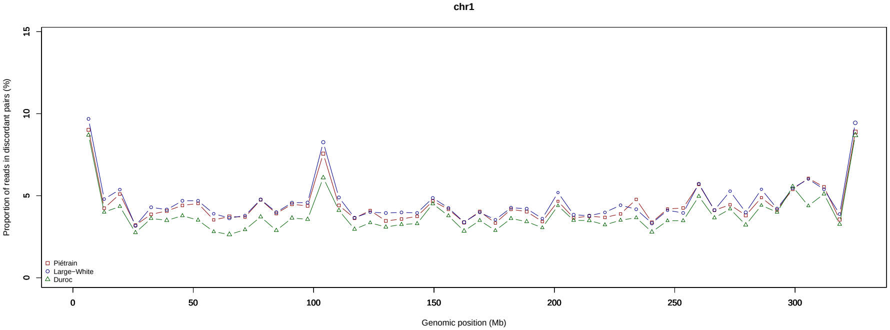
Nouvel assemblage





http://narcisse.toulouse.inra.fr

Variants de structure



Perspectives

- Exploiter les mate-pair inter-clones
- Utiliser *a priori* l'information apportée par l'assemblage actuel
- Détecter les variants de structure

Remerciements

- Claire Kuchly et Frédéric Escudé
- Sébastien Letort (validation de l'assemblage)
- Emmanuel Courcelle (Narcisse)